AIRLINE PASSENGERS' ONLINE SEARCH AND PURCHASE BEHAVIORS

A Thesis
Presented to
The Academic Faculty

by

Misuk Lee

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2009

AIRLINE PASSENGERS' ONLINE SEARCH AND PURCHASE BEHAVIORS

Approved by:

Professor David Goldsman,
Committee Chair
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Laurie Garrow, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Professor Marco Castillo
School of Public Policy
Georgia Institute of Technology

Professor Chip White
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Paul Griffin
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: 1 July 2009

To my husband Chul Jong

ACKNOWLEDGEMENTS

There is not enough space in this document to thank my Ph.D. committee chairs, advisors, and mentors Prof. Laurie Garrow and Prof. Marco Castillo. They gave me the opportunity to earn my Ph.D. by taking a big chance on me. Their advice and friendship has been invaluable to me throughout this project. Laurie is one of the most generous and compassionate people I know, never asking for anything in return.

Above all, I want to thank my parents, Hwa Bun Jang and Won Hwan Lee, who have always trusted and supported me; my husband, Chul Jong, whose love has been fulfilling me "Every step of the way".

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This paper studies airline customers' online search and purchase behaviors. Two fundamental aspects of online behavior are examined: (1) the link between search behavior and buying behavior and (2) the evolution of inter-temporal search and purchase decisions of strategic buyers.

In the first study, we examine online customers' dynamic conversion behaviors using clickstream data. A new model based on Markov chains that incorporates discrete choices and decision-timing is proposed to capture key search effects on consumer decisions as well as dynamics of browsing behavior both within and across visits. Empirical results show that within-site search activities lead to strong consumer engagement and thus increase purchase and revisit propensities. Fit comparison between first and second order Markov chains allows us to conclude that consumer decisions are primarily influenced by the current search. Furthermore, we observe that consumers dynamically adjust their browsing behavior both within and across visits.

The second study investigates the evolution of inter-temporal search and purchase decisions of strategic buyers. Risk neutral buyers follow simple behavioral rules based on future and current prices and options available. We show that the trade-off between waiting and purchasing will become less and less favorable to waiting. Price elasticity should therefore drop as departure date approaches. With stationary price distributions, search and purchase efforts increase with proximity to the deadline. We extend the base model to allow for price evolution and demand uncertainty. We find that increases in mean price and price dispersion may attenuate increasing propensities for search and purchase. We demonstrate our models through a logit estimation on a unique data set from a major online travel agency.

CHAPTER I

INTRODUCTION

In recent years, airlines have faced increasing revenue pressures due to the increased market penetration of low cost carriers, and the prevalence of internet sales. In this competitive environment, some legacy airliners have begun to investigate the viability of offering discount products targeted to highly flexible travelers. The first study uses data from a website that offers customers discount on a known airline and market in exchange for travel date and flight uncertainty. Specifically, the ability to track the sequence of screens viewed by customers, as well as identify repeat visitors provides a unique opportunity to model search and purchase behavior of time-flexible travelers. This paper builds on prior work in the area of airline search and purchase behaviors by (1) proposing the search effects within and across visits on consumer engagement (2) relating the consumer stickiness to a series of consumer decisions on page request, purchase and revisit, and (3) providing a comprehensive view on consumer decisions and decision cycles.

We conceptualize consumer navigation within the site as a series of the following decisions: (1) whether to request an additional page or to exit the site (exit decision), (2) whether to purchase once ending the search (purchase decision), and (3) whether to come back after ending the search without purchase (revisit decision). We associate these decisions with the differential effects of search activities within and across visits. We estimate a discrete choice model of consumer decisions on the clickstream data. Our modeling results verify the key effects on the consumer decisions and provide several findings about browsing behavior. Furthermore, discrete consumer decisions are incorporated with proportional hazard models of decision-timing through a continuous Markov chain model, which provides a comprehensive and simultaneous perspective on decisions and durations.

Key findings fall into one of two clusters: key search effects on consumer decisions and dynamics of browsing behavior both within and across visits. First, our results indicate that page request and purchase decisions are driven by search depth and long page-view time. Moreover, repeated visits and short inter-visit durations increase purchase and revisit propensities. Along with these within-site search effects, higher price discounts lead to strong consumer engagement. In addition, we have verified that price effects are exogenous, that is, the price offers consumers receive are not affected by search depth and durations. Fit comparison between first and second order continuous Markov chain models allows us to conclude that consumer choices are primarily influenced by the current search, independent of the past searches.

Next, we find that consumers dynamically adjust their browsing behavior both within and across visits. Among our findings in browsing dynamics, there first is evidence for time constraints or learning effects in both browsing components (page requests and page-view duration) as consumers go deeper into the site. Second, our results are consistent with the presence of two distinct categories of consumer behaviors: goal-oriented and experiential consumptions that spill over multiple visits. Consumers who revisit after relatively short durations are likely to be more goal-oriented than consumers who revisit after relatively longer intervals. Moreover, goal-oriented consumer behaviors lead to, if any occur, immediate purchase and stronger engagement within the site.

In the second study, we investigate search and purchase behavior in market with perishable goods. Markets of this kind include hotel rooms, newspapers, airline tickets and sports and artistic events. The necessity of search due to costly information acquisition have shown to be fundamental in explaining price dispersion and non-competitive pricing as well as unemployment and real effects of monetary policy. This paper exploits a unique data set on search and purchase behavior from an internet search engine.

Our investigation allows us to test some implications of standard optimal stopping problems and discuss some limitations of the current literature on revenue management. In particular, we show that given the variability of prices, engaging in search is optimal. Or equivalently, that there are strategic reasons to continue searching for a good price. A basic property of the model, however, is that under mild conditions on preferences the trade-off between waiting and purchasing will become less and less favorable to waiting. Price elasticity should therefore drop as departure date approaches. Consistent with this, we find that demand becomes more inelastic as deadline approaches.

Recently Horner and Samuelson (2008) have shown that a monopolist willing to sell a perishable good to a buyer with unknown valuation of the good might engage in dynamic price discrimination. I.e., monopolists without commitment capacity might still extract some rents from consumers. The main conclusion of the paper, which resembles the results on the revenue management literature (Gallego and Ryzin (1993)), is that the optimal sequence of prices will be a decreasing one. This is so because monopolists become more and more pessimistic as time passes since buyers with higher valuations will likely buy early. The result from Horner and Samuelson (2008) is important because it shows that the basic result of the revenue management literature holds even if buyers can search or wait for a better price. The model we develop in this paper has the same property, higher valuation buyers tend to leave the market early.

While the result that the optimal policy for a monopolist is to successively shave prices as deadline approaches is robust to search behavior, it contradicts observed empirical evidence. Prices and its dispersion, in the airline industry, tend to rise as deadline approaches. Consumers would therefore increase their search and purchase as deadline approaches. The data reveal that search does increase as deadline approaches but that purchases given searches decline. Since search is an expression of the gains from searching or the potential benefits of purchasing, this behavior is only explicated by the presence of a small probability of finding unusually low prices. This would simultaneously explain the increase in the expected gain from trade together with the lower success ratio after search.

In order the reconcile the behavior in prices and search and purchase, we extend the basic search model to include demand uncertainty. As shown by Dana (1998), demand uncertainty justifies the use of price sequences that are increasing as deadline approaches and also predicts that unusually low prices might occur with small probability. Theoretically, the inclusion of demand uncertainty do not change the basic properties of the model since demand uncertainty basically means left-ward shift in the distribution of consumer valuation in earlier periods. If anything, demand uncertainty heightens the fact that demand will become more price inelastic as deadline approaches.

CHAPTER II

MARKOV CHAIN MODEL FOR DYNAMIC SEARCH AND PURCHASE BEHAVIORS

2.1 Theory

We seek to model search and purchase decisions in an e-commerce site. The decision variable of interest is (1) whether to request an additional page or to exit the site, (2) whether to purchase once ending the search, and (3) whether to revisit after leaving without purchase. We build a dynamic model that attempts to capture the differential effects of navigation activities on consumer decisions. The key effects we test in our model are introduced below.

2.1.1 Key Effects

2.1.1.1 Price Effects (Exogenous Factor)

The most prominent effect on consumer decisions– especially on purchasing– is price. We investigate not only its direct correlation with purchase decisions but also the overall effects on within-site stickiness. We expect that the lower the offered prices are, the more engaged consumers are in the site. Strong consumer engagement, of course, is more likely to end up with purchasing. An important issue associated with price effects is whether they are exogenous. One might argue that as consumers are more involved in searches, they are more likely to find better offers, i.e. lower prices. If this is the case, causality between price and consumer engagement is not obvious. This paper tests for such endogenous relationships.

While measuring the price effects, we assume that consumers are knowledgeable about other alternatives and use the lowest price offered by other sellers as an anchoring reference price. Consumer decisions are made by assessing prices as discounts or surcharges relative to this reference price. We believe these assumptions are reasonable because: 1) the site where we have collected clickstream data primarily target highly time-flexible and price-sensitive consumers (leisure travelers). 2)The emergence on-line travel agencies and low cost carriers made the comparison of prices across both airline competitors and sales channels much easier and faster. 3) Most important, the use of a reference price for modeling purposes is convenient as we can control for market-specific effects (*e.g.*, departure/arrival airports, number of connections, length of connections, flight durations, etc.) as well as demand effects (*e.g.*, seasonality, holidays, etc.). The use of a reference price is also consistent with the methodology used by other researchers, most notably for Travelocity's "Good-Day-to-Buy" sales campaigns (Smith, *et al.*, 2007).

C1. Higher discount rates lead to stronger consumer engagement.

2.1.1.2 Behavioral Effects (Endogenous Factor)

(a) Intra-visit Exposure Effects

Prior research on repetition effects in marketing suggest two different patterns of consumer response to repeated exposures within the same web visit (e.g. see Buchanan and Morrison (1988), Chatterjee *et al.* (2003) and Hoffman and Novak (1996)). The first pattern posits that response probability decreases over time (wear-out). The second response pattern holds that initial response probability may be low, but increases with repetition to a maximum level (wear-out). In the initial wear-in stage, increased response opportunity with each additional searches lead to an increase in affect. Subsequently, satiation (or tedium) leads to wear-out, when each additional exposure after wear-in has a significant negative effect.

We hypothesize that wear-in dominates in on-line search and purchase behaviors so that consumers are even more engaged within the site as they request additional pages and spend more time in the site. The rationale for this follows from the fact that airline passengers (especially leisure travelers) of our interest are price elastic and are willing to navigate web sites to find a better alternative. Moreover, the Internet environment reduces consumer search costs, which makes consumers search more. Thus, we expect that the more pages consumers request and the longer consumers spend in viewing pages, the stronger consumers are engaged in the site.

C2. The positive effect on consumer engagement due to wear-in may dominate over negative effects due to wear-out.

- *C2-1. More page requests reinforce consumer engagement.*
- *C2-2. Longer page view time reinforces consumer engagement.*

(b) Exposure Effects across Visits

Many researchers including Chatterjee *et al.* (2003), Hoffman and Novak (1996), and Moe (2003), discuss two distinct categories of consumption behavior: goal-oriented (directed) and experiential (exploratory) behaviors. We expect that these consumer behaviors are captured through inter-visit activities. Consumers who revisit after relatively short durations are likely to be more goal-oriented than consumers who revisit after relatively longer intervals. Moreover, goal-oriented consumer behaviors lead to, if any occur, immediate purchase and stronger engagement within the site.

C3. Inter-visit behaviors indicate two distinct consumer populations– goal-oriented and experiential consumers.

- *C3-1. Revisit reflects more goal-oriented consumers (stronger consumer* *engagement).*

- *C3-2. Shorter inter-visit time reflects more goal-oriented consumers (stronger consumer engagement).*

2.1.2 Consumer Decisions

We conceptualize customer navigation as a series of the following binary decisions: (1) whether to request an additional page or to exit the site (exit decision), (2) whether to purchase once ending the search (purchase decision), and (3) whether to come back after ending the search without purchase (revisit decision). We can associate these decisions with the differential effects discussed in the previous section. Purchase and revisit decisions are clearly driven by consumer commitment and loyalty whereas an exit decision is an obvious consequence of consumer disengagement. Therefore, we expect same effects on purchase and revisit decisions and the opposite effect in an exit decision.

E. Strong consumer engagement leads to purchase and revisit decisions.

- *E1. Purchase and revisit decisions move in the same direction (stronger consumer engagement).*

- *E2. Exit decisions (disengagement) move in the opposite direction to purchase and revisit.*

Specifically, we anticipate high discount rates lead to consumer engagement, and thus increase purchase and revisit propensities while decreasing exit propensities. Additional page requests and long page-view durations are supposed to have negative impacts on exit decisions and positive impacts on purchase and revisit. The expected correlations between key effects and consumer decisions are listed in Figure 1.

Table 1: Key effects and expected relations to consumer decisions

	Disengagement	Engagement	
Effects	Exit	Purchase	Revisit
C1. Discount Rates	-	+	+
C2-1. Page Requests	-	+	+
C2-2. Page-View Durations	-		
C3-1. Repeated Visits	-	+	+
C3-2. Inter-visit Durations	+	-	-

2.1.3 Dynamics of Browsing Behavior

We theorize that consumer browsing patterns characterized by intra-visit factors (page request, page view durations) and inter-visit factors (repeated visits, inter-visit time) have differentiated effects on consumer engagement and decision. In the analysis of those effects, we face the following questions: Is each effect related with other effects? Does the interaction between effects, if any, reinforce the corresponding effects or attenuate them? How do we interpret the relations? In this section, we discuss the dynamics of consumer navigation patterns.

2.1.3.1 Intra-visit Dynamics

A proper question on intra-visit dynamics is whether search depth reduces the page view durations or not. We may hypothesize two opposite patterns - time constraints (or learning) and involvement. Bucklin *et. al.* (2003) and Mandel and Johnson (2002) point out a change in browsing behavior can be a response to the salience of time constraints, which may vary as a visitor browses the site or may be the result of varying degrees of involvement with the site or with tasks being performed. Johnson, Bellman, and Lohse (2003) study the duration of Web site sessions across multiple visits. Using session durations from a panel of Internet users, Johnson, Bellman, and Lohse test learning phenomenon. Results indicate that visitors spend less time per session the more they visit the site, which in turn suggests that they become more efficient as they return to the site. If there were time constraints on Internet usage at the individual level, we expected search depth to have negative effects on page-view duration. In contrast, if visitors became more involved as they requested more pages, we expected search depth to be associated with longer page-view durations. We investigate which pattern– time constraint or involvement– dominates in consumer navigation behaviors.

B1. If engagement dominates over time-constraint (or learning) in intra-visit behaviors, page-view duration becomes longer as number of pages.

2.1.3.2 Inter-visit Dynamics

We can also hypothesize two opposite browsing patterns that spill over across visits– learning and involvement. When consumers accumulate the knowledge on within-site alternatives as well as the outside options, and can use the knowledge they acquired in one visit for subsequent visits, they may need less time to come back as they repeatedly visit. The inter-visit time may be shortened over repeated visits. In contrast, repeated visits may imply that consumers may require even more time before making decisions, in which case the inter-visit time will increase over repeated visits. Our proposed modeling approach enables us to investigate how within-site browsing behavior changes as consumers return to a site.

B2. If engagement dominates over learning in revisit behaviors, inter-visit duration becomes longer as number of visits.

2.2 Data

The online clickstream data for this research was collected over a two-year period, from August 2004, to June 2006, and targeted to customers traveling on one of 21 routes between New Zealand, Australia, and/or Fiji. The majority of prices shown to customers were between NZ $300 and $500, where $500 was approximately equal to the lowest price offered on a traditional round trip Freedom Air itinerary.

Online clickstream data captures the sequence of web pages viewed by customers.

We define four terms to describe how clickstream data was used for the analysis. A *search* is defined as the combination of two pages: the page in which a customer inputs search parameters and the page in which the price offered to the consumer is displayed. A *visit* is a sequence of pages that a customer requests within a specific time period to search for a single product (defined as an airline route).

An alternative concept to visit which is often used in the analysis of clickstream data is a *session*. It is an identity tag assigned every time a user launches a new browser. However, we have found that many customers launch several browsers almost simultaneously and search back and forth over multiple browsers. In order to eliminate multiple browser effects, we define a visit as a sequence of searches done within a specific time period regardless of the number of browsers launched. We follow previous research and industry practice (see e.g. Bucklin and Sismeiro(2003) and Moe and Fader (2004)) and assume that a page request started a new visit if it was requested after an idle period of at least 30 minutes. Our data show that time between page requests falls either within 5 minutes or in more than 30 minutes.

On the other hand, confining searches in a visit is a fairly narrow approach to consumer behaviors. We observe that a consumer's decision period may span days, weeks, or even months over several repeated visits. Moreover, many researchers including Bucklin and Sismeiro(2003), Chatterjee *et. al.*(2003), Johnson *et al.* (2004), and Moe and Fader (2004), point out the dynamics of consumer searches across visits or sessions. Customers' search and purchase decisions are better characterized as a sequence of searches over possibly repeated visits. We model customers' searches throughout the whole *purchase decision cycle* (search-exit-revisit-purchase), which may span up to 6 months. Our data indicate that 90% of purchase cycles fall within one month.

A *cookie* is used as a proxy for an individual consumer. As noted by Moe and Fader (2004), one limitation typically associated with clickstream data is that it is difficult to obtain characteristics that identify a particular user. In contrast to the U.S. where retailers are reticent to trace cookies stored in customers' computers due to privacy concerns, New Zealand retailers are much more open to tracking cookies. This provides us with the opportunity to use cookies to identify and track individual customers.

In order to clarify these definitions, Table 2.2 and Figure 1 provide an example of the clickstream data used for this analysis. Each row represents a search that contains search parameters (route, time window, prior notice) and search results (price, purchase decision). The customer associated with cookie User1 visits once and leaves the site after two consecutive searches. This customer does not come back thereafter and his/her decision cycle ends up with no purchase.

On the other hand, the customer associated with cookie User2 shows more engaged behaviors. During the first visit, he/she browses three pages and leave the site without purchasing. Then, he/she comes back shortly after the first visit and purchase. The decision cycle of User2 is made after five searches that occur across two visits. In this analysis, we associate customer search behavior with purchase decision cycle and study the sequence of searches within this cycle in order to explore the dynamics of customer search and purchase decisions.

Table 2: Example clickstream data

Date	Cookie	Route	Visit No.	Search No.	Window	Notice	Price	Purchase
08/24/2006 4:10:16 PM	User1	BNEHLZ	1	1	120	30	150	No
08/24/2006 4:10:30 PM	User1	BNEHLZ	1	2	120	20	130	No
08/25/2006 4:09:40 PM	User2	AKLOOL	1	1	30	10	300	No
08/25/2006 4:10:10 PM	User2	AKLOOL	1	2	35	10	280	No
08/25/2006 4:10:24 PM	User2	AKLOOL	1	3	35	7	275	No
08/25/2006 5:10:15 PM	User2	AKLOOL	2	1	30	10	300	No
08/25/2006 5:10:55 PM	User2	AKLOOL	2	2	35	10	280	Yes

Table 3 presents summary statistics for the clickstream data used in this study. Note that most customers visit the site only once. On average, a customer searches three times throughout the entire visits. The purchase conversion rate per decision

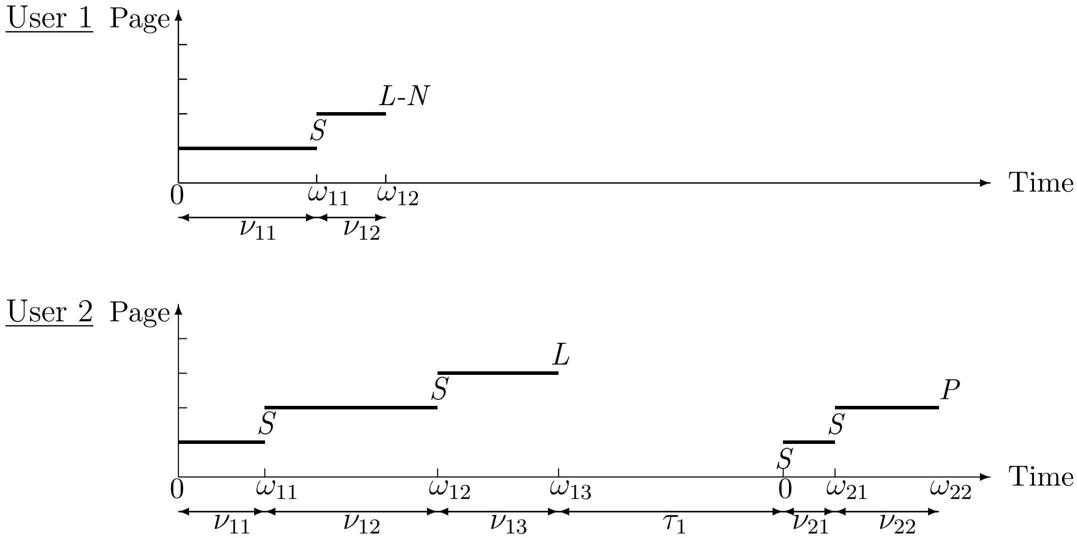

Figure 1: An example of consumer search activities

cycle is approximately 3%. However, customers that ended up purchasing searched, on average, much more than three times.

Table 3: Summary statistics for clickstream data

Data View	Occurrences
Search	42,554
Visit	20,354
Purchase Cycle	15,237
Cookie	12,588
Purchase	474

2.3 Modeling the Consumer Behavior

2.3.1 Customer Utility and Decision

Consumers engage in three sequential stages of binary decisions: (1) whether to request an additional page or to exit the site, (2) whether to purchase once ending the search, and (3) whether to come back after ending the search without purchase. In order to represent the sequence of consumer decisions, we first define customers' navigation states. At any moment, consumers may "Search", "Purchase", be on temporary "Leave" without purchasing, or leave the site forever with "No Purchase".

Starting in state “Search”, a consumer needs to decide whether to search more or not. If he/she decides to request an additional page, his/her next state remains “Search”. Once deciding to end search, he/she decides whether to purchase or just leave. On the former, he/she transits to state “Purchase”. On the latter, he/she move on to state “Leave”. The last decision is whether to come back or not. Consumers in state “Leave” may revisit the site and start again in state “Search”. Figure 2 illustrates how the state changes depending on each decision.

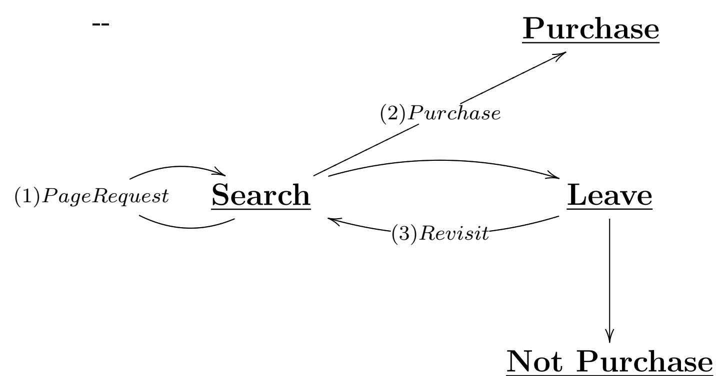

Figure 2: Page-to-Page transitions

We assume that consumer k has latent utility U_{knva} in state a on discrete viewing occasion n of visit v, where there are totals of K consumers, A decisions, V_k visits of consumer k, and N_{kv} viewings for the vth visit of consumer k. Then, we formulate the utility as

$$U_{kvn} = \Gamma_{vn} X_{k,v,n-1} + \epsilon_{kvn}, \quad \epsilon_{kvn} \sim G(0,1). \tag{1}$$

$U_{kvn} = (U_{kvnS}, U_{kvnP}, U_{kvnL}, U_{kvnN})$ is a vector of the utilities associated with four states, “Search”, “Purchase”,“Leave”, and “Not Purchase”.

X_{kvn-1} is a vector of independent variables denoted as

$$X_{kv,n-1} = (1, d_{kv,n-1}, \tau_{kv,n-1}, \nu_{kv,n-1}, \omega_{kv,n-1}, v, n-1)',$$

where $d_{kv,n-1}$ is the discount rate, $\tau_{kv,n-1}$ the inter-arrival time for resuming search(revisit time), $\nu_{kv,n-1}$ the inter-arrival for page-to-page transition, and $\omega_{kv,n-1}$ the cumulative page view time of v-th visit. Note that

$$\omega_{kvn} \quad = \quad \omega_{kv,n-1} + \nu_{kvn}.$$

Γ_{vn} is a 4×6 parameter matrix associated with the dependent variables X_{kvn-1}.

We observe only decisions (equivalently states), not utilities. It is natural to assume that a consumer chooses the state that will bring out maximum utility. Then, we can associate the consumer's selection on the next state with the latent utilities as follows:

$$Y_{kvna} \quad = \quad \begin{cases} 1 & \text{if } U_{kvna} = max_b\{U_{kvnb} I(a(k,v,n-1),b),0\} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$a(k,v,n) \in \{S,P,L,N\}$ is the state of consumer k on page n of visit v, i.e. $a(k,v,n) = b \in A$ such that $Y_{kvnb} = 1$. Matrix $I(i,j)$ indicates that the corresponding transition from i to j is allowed. $I(a(k,v,n-1),b)$ indicates whether the transition from current state $a(k,v,n-1)$ to state b is possible. For example, the transition from "Search" to "Search" is allowed, i.e. $I(S,S) = 1$, but the direct transition from "Search" to "Not Purchase" is not allowed, i.e. $I(S,N) = 0$.

Consumer utilities U_{kvn} are defined as linear functions of search characteristics $X_{k,v,n-1}$ and error terms ϵ_{kvn}. Assuming Gumbel distributed errors, we model the consumer choices Y_{kvna} with hierarchical logit models. By fitting Y_{kvna} on $X_{k,v,n-1}$, we estimate the effects of consumer search characteristics on consumer choices.

2.3.2 Dynamics of Customer Decisions

We have modeled the consumer utilities and the corresponding decisions. Consumer decisions, Y_{kvn}, are based on the current state $a(k,v,n-1)$ and underlying utilities U_{kvn}. Consumer utilities, U_{kvn}, are determined by the current search, $X_{kv,n-1}$. Consequently, the search result $X_{kv,n-1}$ based on the current state $a(k,v,n-1)$ leads to the next decision Y_{kvn}, which, in fact, indicates the next state $a(k, v, n)$.

If we redefine the consumer search state as $(X_{kv,n}, a(k, v, n))$, the next state depends only on the present state, not on the past states. Thus, we may define search and purchase decisions as Markov chains. To see this, let us decompose $(X_{kv,n}, a(k, v, n))$. The present search result $X_{kv,n-1}$ is characterized by a discount rate, search depth (pages, visits), and time between searches (intra- and inter- visit durations). Search depth and $a(k, v, n)$ define the state where customer k is on the discrete occasions. The intra- and inter- visit durations characterize inter-arrival times of consumers' decision transitions.

Specifically, let $M_k(t, d_k)$ denote a Markov chain corresponding to the consumer k's state at time t parameterized by discount rate d_k. As other continuous Markov chains, $M_k(t, d_k)$ is characterized by (1)state space, (2) transition probabilities, and (3) inter-arrival times.

2.3.2.1 Page-to-page Transitions

The discrete state of the Markov chain is search depth and search state, i.e.,(v, n, a). Transitions between states are derived from three sequential stages of binary decisions: (1) whether to request an additional page or to exit the site, (2) whether to purchase once ending the search, and (3) whether to come back after ending the search without purchase. For example, if a consumer requests an additional page, the number of page views increases by one with the state unchanged ("Search"). When a consumer comes back later, the number of visits increases by one and the consumer transits from state

"Leave" to "Search". We identify possible transitions as

$$\begin{aligned}
\text{Page request} &: (v,n,S) \to (v,n+1,S) \\
\text{Purchase} &: (v,n,S) \to (v,n,P) \\
\text{Exit without purchase} &: (v,n,S) \to (v,n,L) \\
\text{No purchase} &: (v,n,L) \to (v,n,N) \\
\text{Revisit} &: (v,n,L) \to (v+1,1,S).
\end{aligned}$$

The discrete transition probabilities follow the above transitions. First, let us define

- f_k : probability of ending search
- p_k : probability of purchase conditional on ending search
- b_k : probability of revisiting.

Then, we can find the probabilities associated with the above possible transitions:

$$\begin{aligned}
P\{M_k = (v,n+1,S)|M_k = (v,n,S)\} &= 1 - f_k \\
P\{M_k = (v,n+1,P)|M_k = (v,n,S)\} &= f_k p_k \\
P\{M_k = (v,n,L)|M_k = (v,n,S)\} &= f_k(1-p_k) \\
P\{M_k = (v,n,N)|M_k = (v,n,L)\} &= 1 - b_k \\
P\{M_k = (v+1,1,S)|M_k = (v,n,L)\} &= b_k.
\end{aligned}$$

2.3.2.2 Time Durations

There are two types of time durations between transitions. Page view time, ν_{kvn}, is corresponding to transitions from *Search* state. Next, inter-arrival times for resuming search (Revisit time), τ_{kvn}, are associated with *Leave* state.

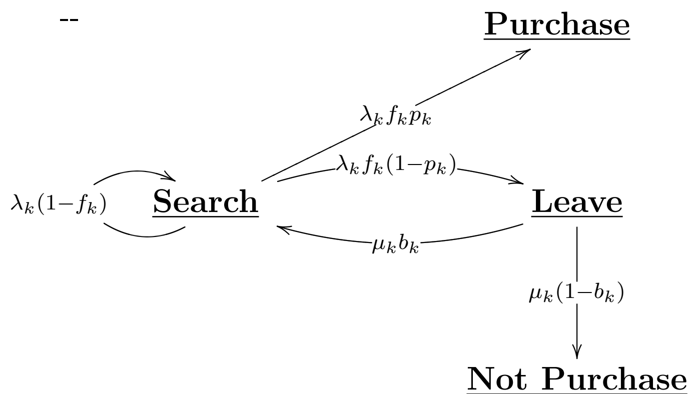

Figure 3: Page-to-Page transitions and transition rates

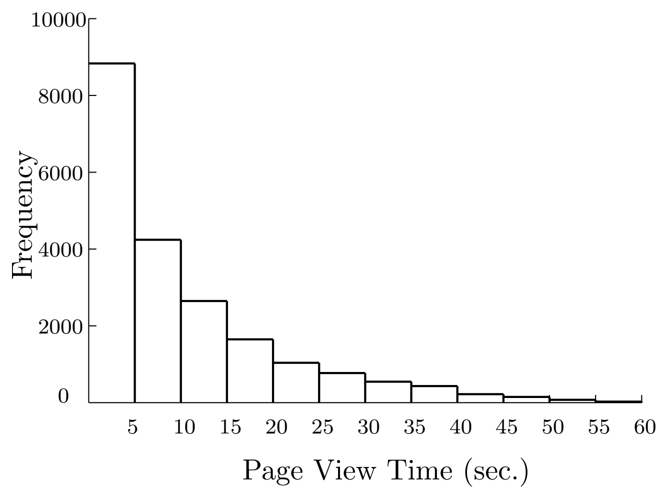

Figure 4: Distribution of page view durations

- Page View Time

  Figure 4 shows the distribution of page-view durations. We observe that majority of consumers spend less than 5 seconds in viewing each page and the frequency of consumers decreases over page view time. We model conditional page view time with state-dependent heterogeneous exponential distributions. Indeed, a large body of literature on purchase-timing decisions in recent years has used the proportional hazard model, which is based on exponential distribution families (see e.g. Allenby *et al.* (1999), Jain and Vilcassim (1991),

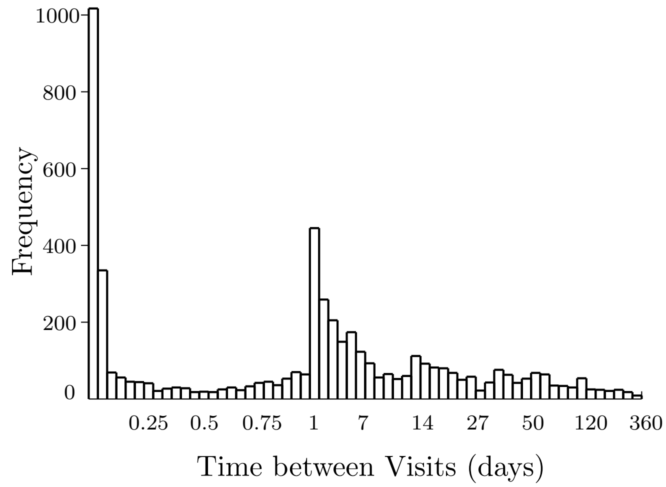

Figure 5: Distribution of inter-visit durations

and Seetharaman and Chintagunta (2003)). We define the probability that consumer k spends more than t at each page as

$$P\{M_k(t+s) = (v,n,S)|M_k(t) = (v,n,S)\} \quad = \quad e^{-\lambda_k t}.$$

This probability is conditional that consumer k decided whether to request additional pages or not. Then, integrating the conditional page-view duration with consumer decisions, we find the unconditional transition rates between states. As illustrated in Figure 2.3.2.1, consumer k decides to keep searching with rate $\lambda_k(1-f_k)$, to purchase with rate $\lambda_k f_k p_k$, or to leave with rate $\lambda f_k(1-p_k)$. Equivalently, unconditional purchase cycle is exponentially distributed with rate $\lambda_k f_k p_k$ and total duration within a visit is also exponential with rate $\lambda_k f_k(1-p_k)$. Note that transition probabilities f_k, p_k are dependent on each consumer's search characteristics, so are the unconditional (page request - purchase - revisit) decision cycles.

On the one hand, we may assume that the rate for page-view durations, $\lambda_k = \lambda$, is constant. We can find the maximum likelihood estimator of the constant rate

λ as follows:

$$\hat{\lambda} = \frac{\sum_{k=1}^{K}\sum_{v=1}^{V_k} N_{kv}}{\sum_{k=1}^{K}\sum_{v=1}^{V_k}\sum_{n=1}^{N_{kv}} \nu_{kvn}}$$

On the other hand, consumers can change the way they browse a site as they request and view additional pages. We investigate the effects of search depth on page-view durations. If there were time constraints on Internet usage, we expected search depth to have negative effects on page-view duration. Learning effects also may cause negative relations between search depth and page-view durations. In contrast, if visitors became more involved as they requested more pages, we expected search depth to be associated with longer page-view time. Therefore, we may extend the model to allow the inter-arrival rate $\lambda_k = \lambda_{vn}$ to depend on the number of page views n and visits v.

- Revisit Time

Next, consider the time until a consumer decides to come back after leaving the site. Figure 5 implies that we may model the conditional inter-arrival time for consumer k to revisit with exponential distributions. As in the case of page-view time, first we may suppose that the inter-arrival rate for revisit is constant, i.e. $\mu_k = \mu$. We can find the maximum likelihood estimator of the constant rate μ:

$$\hat{\mu} = \frac{\sum_{k=1}^{K}\sum_{v=1}^{V_k} N_{kv}}{\sum_{k=1}^{K}\sum_{v=1}^{V_k}\sum_{n=1}^{N_{kv}} \tau_{kvn}}$$

Also, we investigate the effects of repeat visits on inter-visit time by modeling the inter-visit rate as dependent on the number of visits, i.e. $\mu_k = \mu_v$

Repeat visitors may have different motivations to come back. Some of them are highly goal-oriented with an immediate purchase in mind. Some are building knowledge and may need longer inter-visit durations. Figure 5 suggests that inter-visit durations are very disperse, ranging from less than one hour to six months. We therefore extend the model to allow heterogeneous arrival rates across consumers. We suppose that μ_k is Gamma-distributed with parameter θ and δ. Then, the distribution of any inter-arrival time τ is

$$\begin{aligned} P\{\tau_k < t|\mu_k\} &= 1 - e^{-\mu_k t} \\ P\{\tau < t\} &= \int_{-\infty}^{\infty} \left(1 - e^{-\mu_k t}\right) dF(\mu_k) \\ &= 1 - \left(t/(t+\theta)\right)^{-\delta}. \end{aligned}$$

From the above distribution, we obtain the maximum likelihood estimators for θ and δ. The resulting likelihood function is written as

$$L = \prod_{k=1}^{K} \prod_{v=1}^{V_k} \left[\left(\frac{\tau_{kv}}{\tau_{kv} + \theta} \right)^{-\delta-1} \left(\frac{\delta\theta}{(\tau_{kv} + \theta)^2} \right) \right].$$

On the other hand, many researchers including Chatterjee *et al.* (2003), Hoffman and Novak (1996), and Moe (2003), discuss two distinct categories of consumption behavior: goal-oriented (directed) and experiential (exploratory) consumer behaviors. Consumers who revisit after relatively short durations are likely to be more goal-oriented than consumers who revisit after relatively longer intervals. Moreover, goal-oriented consumer behaviors result in, if any occur, immediate purchase. Some highly-engaged consumers may need just a few minutes or hours to make sure that price offer is good enough, or simply to get a credit card. Others may need much more time to explore other alternatives. We then assume two consumer populations: with short inter-visit time and long inter-visit time. Specifically, with probability α, consumer k is a short-duration type, whose inter-arrival rate is gamma distributed with parameter θ_s and δ_s. Other consumers have long revisit time, of which inter-arrival rate is gamma distributed with θ_l and δ_l.

$$\mu_k \sim \begin{cases} Gamma(\theta_s, \delta_s) & \text{w.p. } \alpha_s \\ Gamma(\theta_l, \delta_l) & \text{w.p. } 1 - \alpha_s \end{cases}$$

Then, the distribution of any inter-arrival time follows as

$$P\{\tau < t\} = \alpha_s\Big(1 - \big(t/(t+\theta_s)\big)^{-\delta_s}\Big) + (1-\alpha_s)\Big(1 - \big(t/(t+\theta_l)\big)^{-\delta_l}\Big).$$

The maximum likelihood estimators can be found from the following likelihood function

$$\begin{aligned} L &= \prod_{k=1}^{K}\prod_{v=1}^{V_k}\left[\alpha_s\left(\frac{\tau_{kv}}{\tau_{kv}+\theta_s}\right)^{-\delta_s-1}\left(\frac{\delta_s\theta_s}{(\tau_{kv}+\theta_s)^2}\right)\right. \\ &\qquad \left. +(1-\alpha_s)\left(\frac{\tau_{kv}}{\tau_{kv}+\theta_l}\right)^{-\delta_l-1}\left(\frac{\delta_l\theta_l}{(\tau_{kv}+\theta_l)^2}\right)\right]. \end{aligned}$$

These revisit rates are conditional that consumer k decided whether to come back or not. Therefore, as in models of page view rates, we find the unconditional transition rates after leaving the site, which is illustrated in Figure 2.3.2.1. Unconditional revisit cycle is exponentially distributed with rate $\mu_k b_k$. Again, notice that transition probability b_k is dependent on each consumer's search characteristics, so is the unconditional revisit cycle.

2.4 Results

We apply Markov chain model to the search behavior of our sample of web site visitors. The transition probabilities (or consumers' binary decisions) are estimated using hierarchical logit models. For inter-arrival times between decisions, we estimate the exponential rates. We now present the model results.

2.4.1 Customer Decisions

We apply logit models to three stages of consumer binary decisions (Exit-Purchase-Revisit) and find transition probabilities associated with the three decisions, f_k, p_k, and b_k. Table 4 reports the analysis of consumers' decisions using Equations (1) and (2). Parameter estimates for Γ_{vn} and p-values are given for independent variables of X_{kvn}. The parameter estimates of the proposed model are significant and in the expected direction (see Table 1). While fitting the model, we have used evenly balanced samples. For example, to predict p_k, we have random-sampled equal numbers of purchasing and non-purchasing observations - with the equal sample size of 332 each. Then, taking evenly balanced and large samples into account, we may say reported Pseudo-R^2's reflect a good model fit overall.

Table 4: Transition probabilities of consumer decisions: Regression coefficients

Effects	Variables	f_k	p_k	b_k
Price	d	-.346	1.441	.879
		(.000)	(.000)	(.000)
Inter-visit Durations	τ	.002	-.0112	
		(.000)	(.038)	
Page-view Durations	ν	-2621.080		
		(.000)		
	ω	8254.683		
		(.000)		
Repeated Visits	$1_{\{v=1\}}$	-.760	1.516	1.295
		(.000)	(.000)	(.000)
	$1_{\{v=2\}}$	-.786	2.519	2.159
		(.000)	(.000)	(.000)
	$1_{\{v=3\}}$	-.832	1.792	1.923
		(.000)	(.000)	(.000)
	$1_{\{v>3\}}$	-.816	2.108	3.168
		(.000)	(.000)	(.000)
Page Requests	n	-2.407^{*}	.124	.187
		(.000)	(.002)	(.000)
	constant	.874	-1.783	-1.396
		(.000)	(.000)	(.000)
	Observations	29802	664	7162
	LL	-17063.912	-369.314	-3959.228
	Pseudo R^2	0.173	0.198	0.203

Table 5 reports the fit statistics for three transition probabilities. Notice that the hit rates are more than 60% in all transition probabilities. More than 60% of hit rates are considered as a significant improvement in 50-50 chances of binary choices. As a measure of model adequacy, we have computed out-of-sample predictive performances. We observe almost no differences in hit rates between in-samples and out-of-samples. For purchase and revisit probabilities (p_k, b_k), out-of-sample hit rates

Table 5: Model fit: Prediction success table for consumer decisions

Prob.		First-order M.C. Hit Rate	First-order M.C. Log Likelihood	Second-order M.C. Hit Rate	Second-order M.C. Log Likelihood	Observations
f_k	In-sample	60.76%	-17063.912	60.91%	-16949.631	29802
	Out-of-sample	60.49%	-7353.521	60.76%	-7355.446	12752
p_k	In-sample	62.62%	-369.314	62.86%	-367.341	664
	Out-of-sample	64.11%	-146.970	64.15%	-151.023	284
b_k	In-sample	62.68%	-3959.228	62.90%	-3940.398	7162
	Out-of-sample	63.03%	-1670.850	63.28%	-1662.035	3072

are slightly higher than in-sample rates. We suspect this is due to skewness toward no-purchase and no-revisit observations. Since majority of observations fall in non-purchasing and non-revisiting searches, the logit models are likely to predict toward those directions. The bias may aggravate with small sample sizes, which is the case of out-of-sample predictions.

While modeling consumer decisions, we have assumed that the corresponding utilities are determined by present search characteristics only (first-order Markov chain model). In order to verify this assumption, we also model second-order Markov chains, where consumer utilities are dependent on the present and previous searches. Table 5 indicates no significant differences in model fit (log likelihood) and prediction performance (out-of-sample hit rate). This supports our assumption that consumer choices are primarily influenced by the current search.

2.4.1.1 Price Effects

Table 4 shows significant relations between discount rates (price) and consumer engagement. As we have expected (see Table 1), higher discount rates involve higher purchase and revisit propensities (engagement) and lower exit probabilities (disengagement). Although the price effects on consumer engagement seem apparent, we still cannot avoid endogeneity in the suggested logit model. One might argue that if high discount rate is a result of strong consumer engagement, i.e. repeated visits and page requests, the price effect (especially on exit decisions) may not be obvious as it seems in the results.

Table 6: Testing for endogeneity of price: Regression results

Effects	Variables	d	p_k
Inter-visit Durations	τ	-.0000454	-.0114
		(.535)	(0.033)
Page-view Durations	ν	49.7505	
		(.000)	
Repeated Visits	$1_{\{v=1\}}$	.0485	2.0998
		(.000)	(0.011)
	$1_{\{v=2\}}$	.0605	3.2440
		(.000)	(0.002)
	$1_{\{v=3\}}$	.0850	2.8223
		(.000)	(0.052)
	$1_{\{v>3\}}$	.1164	3.4945
		(.000)	(0.065)
Page Requests	n	.00294	0.1645
		(.000)	(0.015)
	constant	.3971	2.9692
		(.000)	(0.643)
	d		-10.4617
			(0.514)
	Residual $\widehat{\xi}$		11.880
			(0.458)
	Observations	42554	664
	R^2	0.0223	0.1982

To verify Hypothesis *C1*, we investigate the endogeneity of discount rates. Suppose that discount rates are correlated with other variables as follows:

$$d = \pi_0 + \pi_1\tau + \pi_2\nu + \pi_3 1_{\{v=1\}} + \pi_4 1_{\{v=2\}} + \pi_5 1_{\{v=3\}} + \pi_6 1_{\{v>3\}} + \pi_7 n + \xi \quad (3)$$

Table 6 reports the regression result. Although we may observe that repeated visits and page requests have significant effect on discount rates, very low R^2 implies that we may reject the endogeneity. More specifically, we test for endogeneity following Wooldridge (2005). If the discount rate d is an endogenous variable, for example, in the following regression of purchasing probability p_k, the error term ξ for discount rate d in equation (3) be correlated with ϵ in equation (3). The easiest way to test this is to include ξ as an additional regressor in equation (4). Since ξ is not observed, we may use residuals $\widehat{\xi}$ estimated from equation (3).

$$\begin{aligned} p_k &= \gamma_0 + \gamma_1\tau + \gamma_2\nu + \gamma_3 1_{\{v=1\}} + \gamma_4 1_{\{v=2\}} + \gamma_5 1_{\{v=3\}} + \gamma_6 1_{\{v>3\}} + \gamma_7 n \\ &\quad + \sigma_1 \boldsymbol{d} + \sigma_2 \widehat{\xi} + \epsilon \end{aligned} \tag{4}$$

The last column in Table 6 represents regression of purchase probability p_k on the residuals $\widehat{\xi}$ for discount rates. The p-value (=0.458) for $\widehat{\xi}$ allows us to reject the endogeneity of discount rates. We may conclude that consumer search activities may not have significant effects on discount rates, which means that discount rates may not change over repeated searches. High correlation (=0.74) of discount rates between consecutive searches also supports this idea.

2.4.1.2 Intra-visit Effects

Both page requests n and page-view durations ν significantly impact the probability of exiting the site f_k. Table 4 shows that the coefficients of n and ν are negative, as expected. Consumers develop a tendency to become loyal to the Web site as they search more and spend more time in viewing a page. This positive effect of intra-visit exposures on consumer engagement is also consistent with within-site lock-in suggested by Bucklin and Sismeiro (2003) and Zauberman (2003). Although page-view duration does not have a significant impact on purchase and revisit probabilities, search depth still increases these probabilities. This further supports *C2*, our hypothesis on positive effects of search depth and page-view time on consumer loyalty.

Estimating exit probability f_k, we control for visit duration ω, which refers to the total time spent during a visit. The positive and significant coefficients of visit duration implies that there might be time constrains in browsing behavior. Consumers trade off total time spent and exit decisions. Although time constraints attenuate the “stickiness”, results from purchase and revisit models allow us to support our hypothesis on positive intra-visit effects on consumer engagement.

2.4.1.3 Inter-visit Effects

The positive coefficients of repeated visits on purchase and revisit as well as negative signs on exit indicate that consumers build commitment and loyalty as they frequently visit the site. This result is consistent with a lot of research on off-line and on-line marketing including Bucklin and Sismeiro (2003), Chatterjee *et al.* (2003), and Johnson *et al.* (2003). Table 4 shows that the longer it takes for consumers to come back, the less likely they are to purchase and request additional pages. This result is consistent with our hypothesis *C3* that repeated visits and short inter-visit durations increase purchase and revisit propensities and decrease exit propensities. As Bucklin and Sismeiro (2003), Chatterjee *et al.* (2003), and Moe (2003) have suggested, consumers who revisit after relatively short durations are likely to be more goal-oriented. The fact that a highly goal-directed consumer revisits the site reflects his/her commitment and loyalty.

2.4.2 Time Durations

2.4.2.1 Intra-visit Durations

While predicting exit probability f_k, we have found that total time spent during a visit has positive effects on f_k. Consumer may cope with self-imposed or externally imposed time constraints. Figure 6 also shows that page view time tends to decrease over page requests and visits.

The results from the page view time estimation further support time constraints or learning on internet browsing. Table 7 and Figure 2.4.2.1 report the fit of three different models discussed in Section 2.3.2.2 When we allow λ to depend on n, the number of page views, log likelihood improves, while it seems almost unchanged with repeated visits. Page view time is shortened as consumers request additional pages,

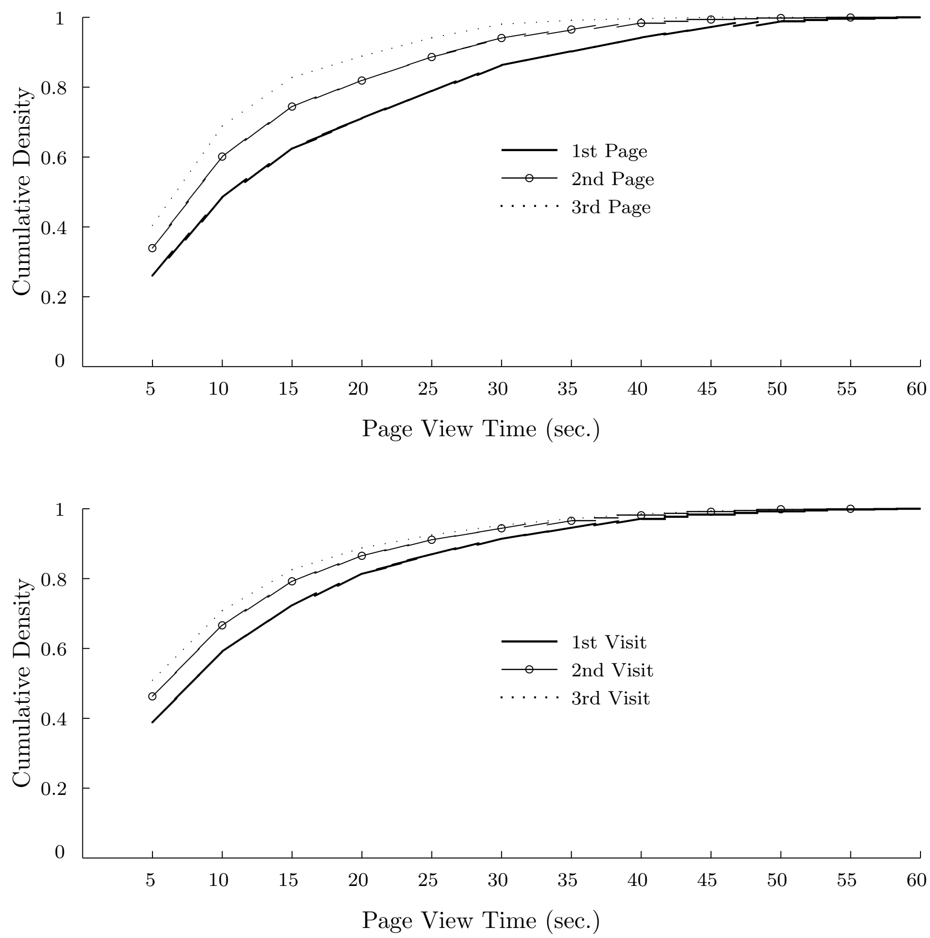

Figure 6: Cumulative distributions of page view times by page requests and visits

but it does not show significant differences over visits. We may conclude that time constraints or learning effects due to repeated searches be valid within a visit.

In the previous section, we have learned that repeated searches and long page view time tend to increase consumer engagement, making consumer request additional pages, purchase, and revisit. These inter-visit exposure effects is attenuated by time constraints or learning as page view time decreases over repeated page requests. Therefore, we may expect inter-visit activities to have even greater impacts on consumer stickiness than they are shown in Table 4.

Table 7: Model fit: Three different models for exponential rates of page view time

Model	Log Likelihood
Constant λ	-16314.6
$\lambda_k = \lambda_n$	-15230.8
$\lambda_k = \lambda_{vn}$	-15219.6
Number of Observations	22000

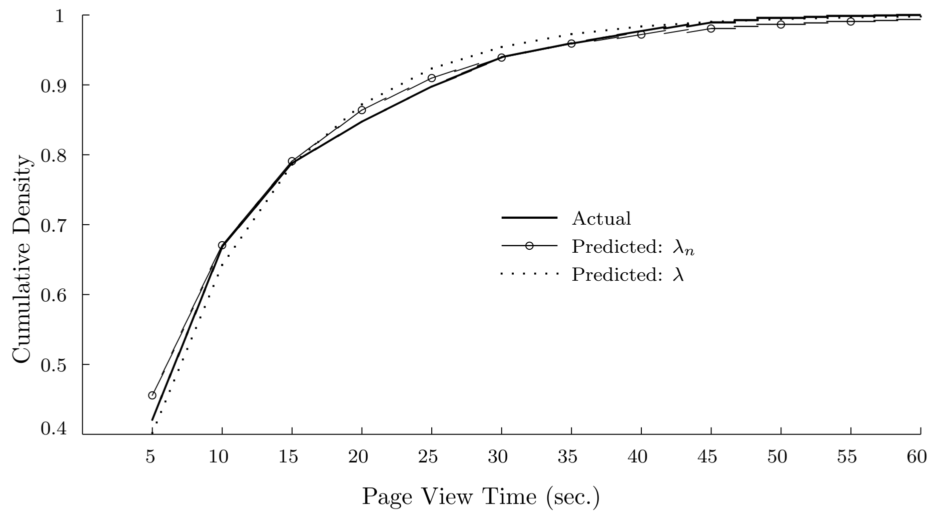

Figure 7: Actual and predicted cumulative distributions of page view time

2.4.2.2 Inter-visit Durations

In section 5.2.2., we have proposed three models for inter-visit rate: (1) constant rate, (2) heterogeneous rate over one consumer population, and (3) two heterogeneous populations. The second model allows consumers to have different motivations to revisit, which are revealed in heterogenous inter-visit rates with Gamma distribution. The third models is based on a large body of marketing literature that explore two distinct categories of consumer behaviors: goal-oriented and experiential consumptions. In this extension, we model inter-visit rates with the composition of two Gamma distribution

We report the fit of three models in Table 8 and Figure 2.4.2.2. We see a large discrepancy between constant rate and heterogeneous rate models. Moreover, results support our assumption on two distinct motivations for revisit.

Table 8: Model fit: Three different models for inter-visit rate

Model	Log Likelihood
Constant μ	-26002.0
$\mu_k \sim Gamma(\theta, \delta)$	-19330.5
$\mu_k \sim \alpha Gamma(\theta_s, \delta_s) + (1-\alpha)Gamma(\theta_l, \delta_l)$	-17936.5
Number of Observations	5117

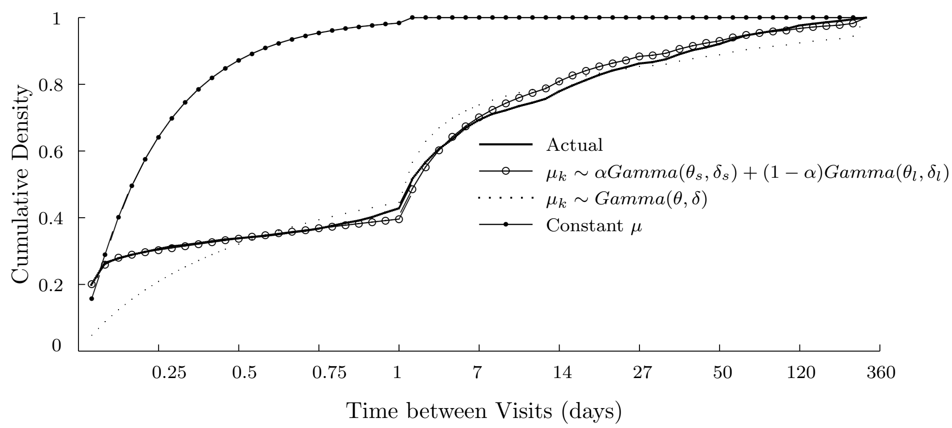

Figure 8: Actual and predicted cumulative distributions of inter-visit time

CHAPTER III

INTER-TEMPORAL SEARCH AND PURCHASE DECISIONS OF STRATEGIC BUYERS

3.1 Search Theory

3.1.1 Model Development

We present the model in the context of a consumer searching for a ticket for a limited period of time. The model considers the decision of how intensely search for a ticket and what ticket to buy given the outcomes of the search. A buyer has K periods to find a ticket and we denote each discrete period by k ($k \leq K$). The intensity of the search in period k is denoted by s_k, where $s_k \in [0, 1]$. At any period, the utility of not having a ticket (or quitting the search process altogether) is denoted $\underline{u}$. The intensity of search, s_k, determines the probability of receiving a quote from the market, p_k, which is the best price available at time k in the market. [1]

In each period, a buyer may have several options to choose from. We assume some distribution F_k over the available options at time k. The consumer incurs cost $-c(s_k)$ if he/she wants to receive a quote with probability s_k. We assume that the cost function is increasing, convex and equals 0 if search is not pursued. If quote p_k is received, the consumer decides to either keep it or to wait until tomorrow and receives expected utility v_{k+1}. If no quote is received, which may occur with probability $(1 - s_k)$, the consumer has no choice but to wait, keeping expected utility v_{k+1}. The decision rule of what quote to keep is simple, buy the best alternative if the expected utility from the alternative is above the expected utility of waiting an

[1]There are other ways to model the search process when several alternatives are available. The main qualitative results of the model are preserved under these alternative hypotheses. We use this model for its simplicity.

extra period. Given this rule and the fact that search intensity does not affect that distribution of alternatives, we can determine the expected utility at time t as follows:

$$v_k(x) \quad = \quad -c(s_k) + s_k E \max\{x - p_k, v_{k+1}(x)\} + (1 - s_k)v_{k+1}(x),$$

where x is the reservation value of the product.

Using some algebra, we can rewrite the original problem as:

$$\begin{aligned} v_k(x) &= -c(s_k) + s_k \int_0^{x - v_{k+1}} (x - p_k - v_{k+1}) dF_k + v_{k+1}(x) \\ &= -c(s_k) + s_k Q_k(x - v_{k+1}) + v_{k+1}(x) \end{aligned}$$

where $Q_k(x - v_{k+1})$ is the average gain from purchasing conditional on searching at time k. Note that $\partial Q_k/\partial x \geq 0; \partial Q_k/\partial v_{k+1} \leq 0$. Since the expected net utility from search should be at least zero, i.e. $-c(s_k) + s_k Q_k(x - v_{k+1}) \geq 0$ for all v_{k+1} and $c(0) = 0$, it must be the case that $v_k \geq v_{k+1}$. This implies that as time progresses, the net utility, $x - v_{k+1}$, increases and therefore the probability of accepting the quote increases. An observable implication of this result is that demand will be more price inelastic as time progresses.

The assumptions made so far already allow us to derive some interesting results if price distributions are stationary and therefore $F_k = F$. The optimal intensity of search should equate the marginal cost of search with the expected marginal benefit of search: $c'(s_t) = \int_0^{x - v_{k+1}} (x - p_k - v_{k+1}) dF_k$. This implies that with the stationary price distributions, the intensity of search should increase with time. To see this, notice that $\int_0^{x - v_k} (x - p_k - v_k) dF_k \leq \int_0^{x - v_{k+1}} (x - p_k - v_{k+1}) dF_k \geq 0$ since the area of integration increases and v_k is monotone decreasing. Since the cost function is increasing in s, we must conclude that $s_k \leq s_{k+1}$.

Using the above consumer utility, we can also derive the purchase probability conditional on search, denoted by bl_k and often called book-to-look ratio in industry practice. A consumer accepts a quote if the net utility from the alternative is greater than the expected utility of waiting another period, which leads to the book-to-look ratio as

$$bl_k(x) \;=\; Pr(x - p_k \geq v_{k+1}(x)) = F_k(x - v_{k+1}).$$

The unconditional purchase probability $b_k(x)$ follows the book-to-look and search intensity, that is,

$$b_k(x) \;=\; s_k(x)Pr(x - p_k \geq v_{k+1}(x)) = s_k(x)F_k(x - v_{k+1}).$$

Taking monotonicity of s_k and v_k into account, we summarize the properties of book-to-look, search intensity, and booking probability as follows:

$$\begin{aligned} F(x - v_{k+1}) = bl_k(x) &> bl_{k-1}(x) = F(x - v_k) \\ c'^{-1}\Big(Q\big(x - v_{k+1}\big)\Big) = s_k(x) &> s_{k-1}(x) = c'^{-1}\Big(Q\big(x - v_k\big)\Big) \\ s_k(x)bl_k(x) = b_k(x) &> b_{k-1}(x) = s_{k-1}(x)bl_{k-1}(x) \end{aligned}$$

Since consumers observe the expected gain from search, $Q_k(x - v_{k+1})$ decreasing over time, they are more likely to search and purchase. Intuitively, as deadline comes closer, consumers feel possible alternatives diminishing and become even more impatient.

Figure 9 represents the average book-to-look ratio and number of searches by weeks from departure. As a whole, book-to-look increases by almost 30%, which implies consumers change the way they search and purchase over time. This supports our assumption that consumer behaviors are strategic in the sense that consumers make decisions based on the evolution of their valuations and prices over time. If consumers are myopic, their decisions are based on the current set of available fares and their valuations. They don't develop expectations on future prices, even without having to wait and possibly purchase in the future. Thus, myopic consumers will not change their purchase behavior over time. While the number of searches almost monotonically increase, book-to-look tend to decrease during the last week. We attempts to capture the differential effects on this phenomenon.

Before going further, we want to point out the relevance of search data. What we observe from data is the number of searches, not search probabilities. Suppose that there are initially n_0 number of consumers in the market, in which we see $n_0 s_1$ number of searches and $n_0 s_1 b l_1$ consumer purchases. Then, in the second period, with $n_0(1 - s_1 b l_1)$ remaining consumers, we observe $n_0(1 - s_1 b l_1)s_2$ search activities. As consumers leave after purchasing over time, the number of searches becomes a complicated function of search intensities and book-to-look's. However, the number of searches has an important implication on search intensities – increasing number of searches means increases in search intensities. To see this, suppose that searches $n_0(1 - s_1 b l_1)s_2$ in the second period is greater than the initial search activities $n_0 s_1$. This is equivalent to

$$n_0(1 - s_1 b l_1)s_2 \geq n_0 s_1 \quad \Longleftrightarrow \quad s_2 \geq \frac{s_1}{(1 - s_1 b l_1)} = s_1,$$

which means increasing search intensity. Since our data shows increasing search activities, we may conclude that consumers search intensities increase as well.

3.1.2 Evolution of Price Distribution

First, we verify the assumption on stationary price distributions. Figure 10 shows that both mean price and price dispersion increase – especially during the last period. We investigate the effects of increasing price dispersion and mean price on search and purchase probabilities. On the other hand, the complete changes in the underlying price distributions discourage us to analyze the price evolution effects. Throughout this section, we assume that the underlying price distribution should be of the same type while changes in mean and dispersion are allowed.

- Mean-preserving spread

The book-to-look $F_k(x - v_{k+1})$ decreases with a mean-preserving spread over F_k. Intuitively, given the same mean price, consumers face more price volatility. Although consumers may have better opportunities to see low prices, at the same time the chances of getting high prices will be even higher. Since purchase decision is determined by the cut-off price that his/her reservation utility $(x - v_{k+1})$ exceeds, book-to-look concerns the upper tail probability or the odds of getting high prices. Therefore, with all things constant, The more volatile prices become, the less likely consumers are to purchase. Furthermore, a mean-preserving spread has same effects on search intensity. All things constant, the average gain from search, $Q_k(x) = \int_0^{x-v_{k+1}}(x-p_k-v_{k+1})dF_k(p_k)$, decreases with a mean-preserving spread over F_k, which reduces search propensities. That is, an increase in the price dispersion discourages not only purchase but also search propensities.

These findings, however, have more complicated effects on inter-temporal search and purchase behaviors. Recall that net utilities $(x - v_{k+1})$ should always increase over time. While the increase in consumer utilities enhances both search and purchase propensities over time, an increase in price volatility attenuates these positive effects.

On the other hand, Figure 9 shows mixed results on search and book-to-look. Until the last period, price dispersions and mean prices seem stationary, in which both search and purchase activities are increasing over time. However, in the last period, the book-to-look decreases whereas the increasing search intensity persists. This phenomenon may happen if the average gain from search $Q_k(x) = \int_0^{x-v_{k+1}}(x - p_k - v_{k+1})dF_k(p_k)$, increases due to increasing net utility $(x-v_{k+1})$ and at the same time, $F_k(x-v_{k+1})$ – the chances of getting low prices – decreases even faster than v_{k+1}. Intuitively, the book-to-look may go down if the tail of F_k is dramatically getting fatter.

The question is whether price volatility can reverse increasing trend of purchase and thus actually reduces the purchase propensities over time. The following simple example of two-period decisions shows that it can occur.

Example 3.1.2.1 Mean-preserving spread

Consider the following price distributions over two periods.

$$
\begin{aligned}
p_1 &= \begin{cases} \underline{p} & \text{w.p. } \beta \\ \overline{p} & \text{w.p. } 1-\beta \end{cases} \\
p_2 &= \begin{cases} \underline{p}-\epsilon & \text{w.p. } \frac{1}{2}\beta \\ \underline{p}+\epsilon & \text{w.p. } \frac{1}{2}\beta \\ \overline{p}-\epsilon & \text{w.p. } \frac{1}{2}(1-\beta) \\ \overline{p}+\epsilon & \text{w.p. } \frac{1}{2}(1-\beta) \end{cases}
\end{aligned}
$$

where $0 < \underline{p}-\epsilon < \underline{p} < x < \underline{p}+\epsilon < \overline{p}-\epsilon < \overline{p} < \overline{p}+\epsilon$.

Note that $E[p_1] = E[p_2]$ and $Var[p_1] + \epsilon^2 = Var[p_2]$. Consumers' search cost function is defined as $c(s) = c^2$.

Then, the search and the book-to-look probability in each period are found as

$$
\begin{aligned}
v_1 &= \frac{\beta^2}{4}\left((x-\underline{p}) - \left(\frac{\beta(x-(\underline{p}-\epsilon))}{4}\right)^2\right)^2 + v_2 \geq v_2 \\
s_1 &= \frac{\beta}{2}\left((x-\underline{p}) - \left(\frac{\beta(x-(\underline{p}-\epsilon))}{4}\right)^2\right) \leq s_2 \\
bl_1 &= P\{x-p_1 > v_2\} = \beta > bl_2
\end{aligned}
$$

and

$$
\begin{aligned}
v_2 &= \left(\frac{\beta(x-(\underline{p}-\epsilon))}{4}\right)^2 \\
s_2 &= \frac{\beta(x-(\underline{p}-\epsilon))}{4} \\
bl_2 &= \frac{\beta}{2}
\end{aligned}
$$

In this example, the right tail of prices distributions becomes much fatter. Specifically, the probability of getting high prices increases dramatically from $(1-\beta)$ to $(1-\frac{\beta}{2})$. Thus, the book-to-look declines over time (from β to $\frac{\beta}{2}$) while the average gain from search increases due to reduced reservation utility $v_k \geq v_{k+1}$.

- Increasing mean (with constant dispersion)

  Given the same distribution types and constant volatility in prices, an increase in mean prices has the same effect as a mean-preserving spread. We expect that both book-to-look $F_k(x - v_{k+1})$ and search intensity $c'^{-1}\Big(Q_k\big(x - v_{k+1}\big)\Big)$ diminish as the distribution shifts toward high prices. Thus, the shift of the price distribution attenuates increasing propensities of search and purchase. The following example demonstrates that increasing mean price may induce decreasing book-to-look.

  Example 3.1.2.2 Increasing Mean (with constant dispersion)

$$
\begin{aligned}
p_1 &= \begin{cases} \underline{p} & \text{w.p. } \beta \\ \overline{p} & \text{w.p. } 1-\beta \end{cases} \\
p_2 &= \begin{cases} \underline{p}+\epsilon & \text{w.p. } \beta \\ \overline{p}+\epsilon & \text{w.p. } 1-\beta \end{cases}
\end{aligned}
$$

where $\underline{p} < \underline{p}+\epsilon < \overline{p} < x < \overline{p}+\epsilon \leq 1$.

Consumers' search cost function is defined as $c(s) = c^2$.

$$
\begin{aligned}
v_1 &= \left(\frac{\big((x-\overline{p}-v_2)+\beta(\overline{p}-\underline{p})\big)}{2}\right)^2 + v_2 \\
s_1 &= \left(\frac{\big((x-\overline{p}-v_2)+\beta(\overline{p}-\underline{p})\big)}{2}\right) \\
bl_1 &= 1
\end{aligned}
$$

and

$$
\begin{aligned}
v_2 &= \left(\frac{\beta\big(x-(\underline{p}+\epsilon)\big)}{2}\right)^2 \\
s_2 &= \frac{\beta\big(x-(\underline{p}+\epsilon)\big)}{2} \\
bl_2 &= \beta.
\end{aligned}
$$

We see the book-to-look decreasing over time. Furthermore, with specific values of $\overline{p}, \underline{p}, x, \epsilon$ and β, search intensity may increase, i.e. $s_1 \leq s_2$ (see Appendix).

3.1.3 Demand Uncertainty

We have learned that the evolution of price distributions may contribute to declining book-to-look. In this section, we consider demand uncertainty to explain the mixed results on inter-temporal search and purchase behaviors. Let us define demand certainty, α, as the probability that a consumer actually ends up with needing the product. We start our analysis with a two-period problem. Indeed, our data show that the changes in consumers' search and purchase behaviors occur in the last period, which implies that our assumption on two-periods may be sufficient. We can solve consumers' decision problems backward.

- Last period ($k = 2$)

  Consumers, offered p_2, make decisions on whether to purchase or not. At this time, they knows whether they will actually need the product. That is, the demand uncertainty is resolved. Without loss of generality, assume that the utility associated with no purchase is zero.

  In the last period, if they didn't purchase in the advance period, they will have to purchase the product with probability α. Then, consumers need to maximize their expected utility over their search efforts.

$$\begin{aligned}
v_2(x) &= \max_s\{-c(s) + sE_{p_2}[\max\{(x-p_2),0\}] + (1-s)\cdot 0\} \\
&= \max_s\{-c(s) + s\big(\int_0^x (x-p_2)dF_2(p_2)\big)\} \\
&= \max_s\{-c(s) + sQ_2(x)\}
\end{aligned}$$

Remember that $\frac{\partial Q_k(a)}{\partial a} = F_k(a)$.

$$\begin{aligned}
v_2(x) &= -c(s_2) + s_2 Q_2\big(x\big) \\
s_2(x) &= c_1^{-1}\Big(Q_2\big(x\big)\Big) \\
b_2(x) &= s_2(x)F_2(x) = c_1^{-1}\Big(Q_2\big(x\big)\Big)F_2(x) \\
bl_2(x) &= F_2(x)
\end{aligned}$$

- Advance period ($k = 1$)

  If they purchase in this period (advance purchase), their expected utility is

$$(x - p_1)\alpha - p_1(1-\alpha) = \alpha x - p_1.$$

  If they don't purchase, the expected utility is just $\alpha v_2(x, \alpha)$. Since they simply procrastinate, they get the utility of waiting another period, $\alpha v_2(x, \alpha)$, in which the outcome of future decisions are taken into account.

  Considering two options, consumers need to maximize their expected utility over their search efforts.

$$\begin{aligned}
v_1(x,\alpha) &= \max_s\{-c(s) + sE_{p_1}[\max\{\alpha x - p_1, \alpha v_2(x)\}] + (1-s)\alpha v_2(x)\} \\
&= \max_s\{-c(s) + sQ_1(\alpha x - \alpha v_2(x)) + \alpha v_2(x)\}
\end{aligned}$$

$$\begin{aligned}
v_1(x,\alpha) &= -c(s_1) + s_1 Q_1\big(\alpha x - \alpha v_2\big) + \alpha v_2 \\
s_1(x,\alpha) &= c_1^{-1}\Big(Q_1\big(\alpha x - \alpha v_2\big)\Big) \\
b_1(x,\alpha) &= s_1(x,\alpha)F_1(\alpha x - \alpha v_2) = c_1^{-1}\Big(Q_1\big(\alpha x - \alpha v_2\big)\Big)F_1(\alpha x - \alpha v_2) \\
bl_1(x,\alpha) &= F_1(\alpha x - \alpha v_2)
\end{aligned}$$

A stationarity assumption on price distributions allow us further intuitions. That is, let us assume that $F_1(a) = F_2(a), \forall a$. Note that $v_2(x) \geq 0$ and $Q_k''(a) = f_k(a) > 0, \forall a$, which means Q_k is convex. Then, we have the following monotone behaviors:

$$\begin{aligned} F_2(x) &> F_1(\alpha x - \alpha v_2) \\ c_1^{-1}\Big(\alpha Q_2(x)\Big) = s_2(x) &> s_1(x,\alpha) = c_1^{-1}\Big(Q_1(\alpha x - \alpha v_2)\Big) \\ s_2(x)F_2(x) = b_2(x) &> b_1(x,\alpha) = s_1(x,\alpha)F_1(\alpha x - \alpha v_2) \end{aligned}$$

, which implies that book-to-look, search, and purchase propensities all increase.

Even without demand uncertainty, if prices are stationary, we expect higher search and book-to-look propensities over time. Demand uncertainty enlarges the gap between two periods. While the unresolved uncertainty in the advance period discourages search and purchase, in the last period, consumers with certain demand only remain and search even more intensively. We notice that under demand uncertainty, the actual number of purchases – not book-to-look – may go down in the last period as consumers without having to purchase have already left. On the other hand, demand uncertainty, if incorporated with heterogenous consumer valuation, may make average book-to-look decline over time.

Let us consider heterogenous valuation over consumers. For simplicity, we assume that there are two consumer populations, for example, business and leisure travelers, with different reservation valuations – high x_h and low x_l valuation . Furthermore, their demand uncertainties are α_h and α_l, respectively. It is reasonable to assume that high value consumers have higher demand uncertainty (smaller α), i.e. $\alpha_h < \alpha_l$. Without loss of generality, let us further assume $\alpha_l = 1$, which implies that low value consumers will always end up with wanting the product.

	High Valuation		Low Valuation	
Period	Population	Book-to-look	Population	Book-to-look
$t=1$	n_h	$F(\alpha_h x_h - \alpha_h v_2(x_h))$	n_l	$F(x_l - v_2(x_l))$
$t=2$	$n_h(1 - s_1(x_h,\alpha_h)bl_1(x_h,\alpha_h))\alpha_h$	$F(x_h)$	$n_l(1 - s_1(x_l)bl_1(x_l))$	$F(x_l)$

Suppose there are initially $0 \leq n_h \leq 1$ high valuation and n_l low valuation consumers. The ratio of two consumers would change at time $t = 2$. High value consumers are more likely to purchase and leave in the advance period, i.e. $s_1(x_h, , \alpha_h)bl_1(x_h, , \alpha_h) > s_1(x_l)bl_1(x_l)$. Moreover, high valuation consumers who don't need the product also leave the market. Therefore, the ratio of high valuation population will drop at time $t = 2$. On the other hand, high value consumers have higher purchase propensities per search at time $t = 2$, i.e. $F(x_h) > F(x_l)$. Consequently, the decrease in the ratio of high value population reduces the average book-to-look of the last period. The larger the difference in the valuations and the demand uncertainty are, the more likely the book-to-look is to decline. The following example demonstrates how the average book-to-look decreases over time with demand uncertainty and heterogenous valuations.

Example 3.1.3. Demand uncertainty and heterogeneous consumer valuation (with stationary price distributions)

Suppose that price distributions are stationary:

$$p_1 = p_2 = \begin{cases} 0.2 & \text{w.p. } 0.3 \\ 0.3 & \text{w.p. } 0.5 \\ 0.8 & \text{w.p. } 0.2 \end{cases}$$

with

$$x_l = 0.3, \qquad x_h = 1, \quad \alpha_h = 0.4$$
$$n_l = 0.8, \qquad n_h = 0.2$$

Then, we can find consumers' search and purchase propensities as follows (see Appendix for details):

Note that for each consumer population, the book-to-look increases over time. In each period, high valuation consumers show higher search and purchase propensities than low valuation consumers. Although low valuation consumers dominate the

Table 9: An example of inter-temporal search and purchase propensities under demand uncertainty

t	Low Valuation		High Valuation		Average	
1	Population	0.8	Population	0.2		
	v_1^l	0.0045	v_1^h	0.1027		
	s_1^l	0.0151	s_1^h	0.0589	avg. Search prob.	0.0238
	b_1^l	0.0452	b_1^h	0.0472	avg. Purchase prob.	0.0131
	bl_1^l	0.3	bl_1^h	0.8	avg. Book-to-Look	0.4
2	Population	0.796	Population	0.076		
	v_2^l	0.00023	v_2^h	0.099		
	s_2^l	0.015	s_2^h	0.315	avg. Search prob.	0.04121
	b_2^l	0.0045	b_2^h	0.315	avg. Purchase prob.	0.0316
	bl_2^l	0.3	bl_2^h	1	avg. Book-to-Look	0.3611

population over all periods, its proportion becomes even higher. Then, the average book-to-look in the last period is much more dominated by low valuation consumers. As a result, the average book-to-look goes down. On the other hand, the average search and unconditional purchase probabilities rather increases over time because high valuation consumers, though small in number, search even more intensively in the last period and account for much of overall search activities.

3.2 The Airline Ticket Market

In recent years, airlines have faced increasing revenue pressures due to the increased market penetration of low cost carriers, and the prevalence of internet sales. The emergence of online travel agencies (*e.g.*, Expedia, Orbitz, Travelocity, etc.) facilitated the comparison of prices across airline competitors and the ever-increasing penetration of low cost carriers (LCCs) that employ very different pricing models than those used by legacy carriers. Specifically, the majority of LCCs in the U.S. use one-way pricing, which results in separate price quotes for the departing and returning portions of a trip. One-way pricing effectively eliminates the ability to segment business and leisure travelers based on a Saturday night stay requirement (*i.e.*, business travelers are less likely to have a trip that involves a Saturday night stay). Combine the use of one-way pricing with the fact that the internet has increased the transparency of prices for consumers and the result is that today, almost half of all air leisure travelers state that they purchase the lowest price they find when using online channels (Harteveldt, *et al.*, 2004).

In hindsight, it is clear that the Internet has transformed the travel industry. For example, in 2007, approximately 55 million (or one in four) US adults traveled by commercial air and were Internet users (PhoCusWright, 2008). In 2006, more than 365 million US households spent a total of $74.4 billion on leisure travel online (Harteveldt, 2006). In addition, as of 2004, more than half of all leisure travel purchases were made online (Aaron, 2007). In many ways, the internet has been both a blessing and a curse for carriers. On one hand, carriers have benefited from lower distribution costs and the ability to interact directly with consumers (versus relying on an intermediary travel agency). On the other hand, the internet has not only increased the transparency of prices for consumers, but for competitors as well. Monitoring competitive prices and seat availability (a measure of demand on competitors' flights) is becoming more common and viable at a large scale. The net result is a highly competitive market in which the ability to segment customers and price discriminate is becoming more difficult and price changes are quickly matched by competitors.

3.3 Data

The online search and purchase data for this research was collected over a two-month period, from October 15, 2007, to December 15, 2007, and targeted to customers traveling between November 15 and December 15 on one of 55 domestic routes in the United States.

Table 10 presents summary statistics for the data used in this study. The purchase conversion rate per search is approximately 3%. The majority of prices shown to

Table 10: Descriptive statistics

Variables	Value
Observations	29095
Number of Markets	55
Departure Dates	31
Number of Searches	248342
Number of Purchases	8847
Average Price	297
Standard Deviation of Price	206

customers were between \$150 and \$350. The price dispersion is quite high with standard deviation 206.52.

3.4 Results

3.4.1 Basic Results

We first illustrate graphically consumer search and purchase propensities. Figure 9 presents daily average book-to-look ratio and number of searches per route.[2] As deadline comes closer, consumers are more likely to search and purchase overall, which was discussed in the base model. Consumers who become more impatient over time, lower their reservation valuations, i.e. $v_k > v_{k+1}$. This consumer impatience leads to more intensive searches and higher purchase propensities. This trend, however, changes in the last period, as shown in Figure 9. While strong search propensities persist till the last period, purchase propensities go down. We have discussed two possible effects behind this phenomenon – price evolution and demand uncertainty. In reality, demand uncertainty are hardly observable from aggregate data. Then, the discussion will be primarily on price evolution.

Figure 10 and Figure 11 show that both mean and variation of price increase dramatically during the last two weeks. We have learned that increases in mean price and price dispersion have negative effects on purchase propensities. In particular,

[2]Figures are based on sample data. It does not necessarily represent all search activities.

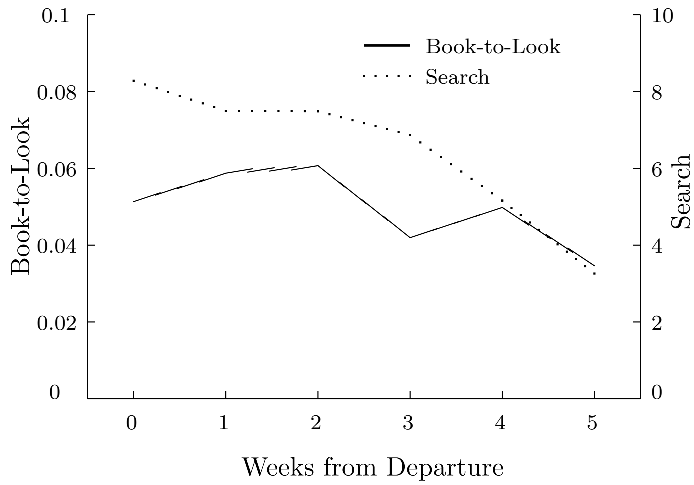

Figure 9: Book-to-look and search

the upper tail of price distribution may enlarge faster than consumers lower their reservation utility. As a result, $F_k(x - v_{k+1})$ – the chances of getting lower prices than the reservation value – may decrease overall.

3.4.2 Estimates

We find the book-to-look at time t as $Pr(x - v_{k+1} - p_k \geq 0)$. Assuming $p_k = \mu_k - \epsilon_k$ and a Gumbel-distributed error ξ_k with mode zero and some scale parameter $1/\beta_k$, we can model the net utility scaled by β_k as follows:

$$\begin{aligned} u_k &= \beta_k(x - v_{k+1} - \mu_k + \xi_k) \\ &= \delta_k - \beta_k \mu_k + \epsilon_k, \quad \epsilon_k \sim G(0,1), \end{aligned}$$

where δ_k is consumer valuation and ϵ_k is a standard Gumbel error. Furthermore, we allow the valuation dependent on the observation i, departure date j, days from departure k, and the market (route) m. This specifies the net utility as

$$u_{ijkm} = \beta_{jkm} x_{ijkm} - \beta_k \mu_k + \epsilon_{ijkm}, \quad \epsilon_{ijkm} \sim G(0,1), \tag{5}$$

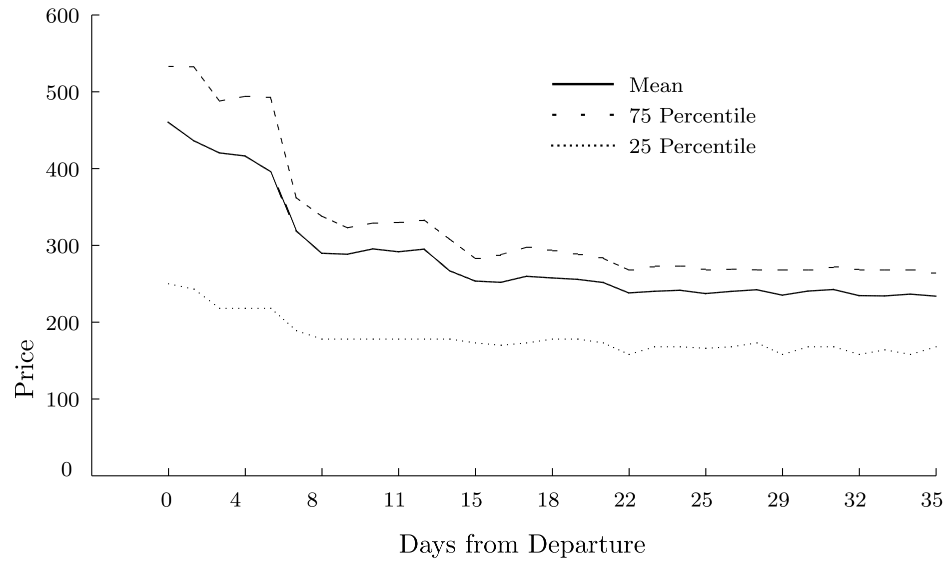

Figure 10: Evolution of mean price and price dispersion

where x_{kjm} is the vector of covariates – dummy variables for market, departure date, and day of week, weeks from departure, and a constant. Based on the above consumer utility, we estimate a logit model on $Pr(x - v_{k+1} - p_k \geq 0)$, the probability of purchase given search controlling for price and price dispersion and a series of dummies to control for the unobserved heterogeneity of markets or routes. The estimation also includes interaction terms of price and weeks to departure.

Table 11 reports the regression results based on equation (1).[3] The first logit model assumes that (Gumbel) price distributions are stationary, i.e. $\mu_k = \mu$, for all k while the second model assumes non-stationary distribution. The log likelihood indicates a better fit of the non-stationary distribution model. Moreover, the constant in the first model is not significant with p-value 0.370 because all the variances due to the changes in price distribution are reflected in the constant. This argues that the evolution of price distribution has a significant effect on consumers' purchase probability.

[3]The dummy variables for market and departure date are not presented. These dummy variables are statistically significant overall.

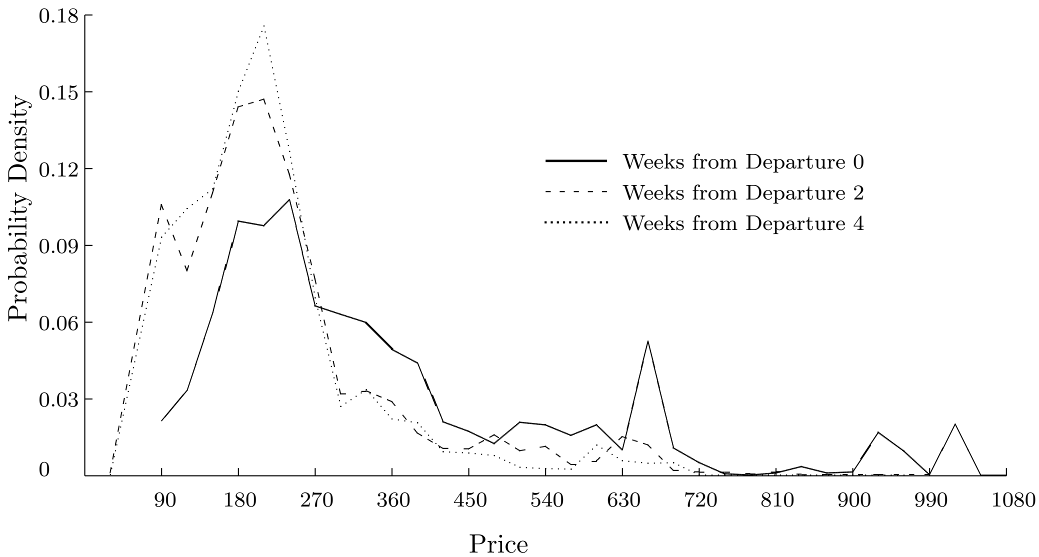

Figure 11: Evolution of price distribution

The estimation shows that unless for the last week of purchase, demand is downward slopping. More importantly, we confirm that demand becomes more inelastic as deadline approaches. This is true in all the specification of the model. Note that this is consistent with our basic model of search but also with the model that has an added demand uncertainty component. Consistent with search theory as well (and risk neutrality), we find that price dispersion delays purchase. This effect can only be explained by the search theoretic reasons.

Model 2 shows that the coefficients on price are significant and decrease overall in absolute terms. That is, demand becomes less elastic to prices. We expect that as consumers lower their reservation values over time $v_k \geq v_{k+1}$, they become more inelastic to price increase. The rate of price inelasticity growth slows in the last period. The increase in mean price and price dispersion reduces the consumer utility, $v_k(x) = -c(s_k) + s_k \int_0^{x-v_{k+1}} (x - p_k - v_{k+1}) dF_k + v_{k+1}(x)$, since $F_k(\cdot) \leq F_{k-1}(\cdot)$. The reduced consumer utility discourages purchasing, which attenuate price inelasticity. This result is consistent with decreasing book-to-look ratio in the last period.

We also estimate our model using the assumption that the error term ϵ_{ijkm} is normal distributed. The result of probit model estimation is compared with logit models in Table 11. The log likelihood and significance of coefficients improve slightly in the probit model. This may imply that price might be modeled with normal distribution, but we cannot conclude the significance of the differences between two models.

Table 11: Regression Results

Variables	Logit 1	Logit 2	Probit
log(price)	-.2441	-.4785	-.2878
	(.000)	(.001)	(.000)
log(stdev(price))	-.0791	-.0471	-.0314
	(.001)	(.061)	(.015)
weeks from departure	.2357	.69732	.4183
	(.000)	(.000)	(.000)
thanksgiving	$-.0943$	-.2001	-.1393
	(.497)	(.156)	(.071)
log(price)$1_{\{w=0\}}$		.3938	.2516
		(.005)	(.000)
log(price)$1_{\{w=1\}}$		.39161	.2402
		(.001)	(.000)
log(price)$1_{\{w=2\}}$		.2305	.1697
		(.000)	(.000)
log(price)$1_{\{w=3\}}$		.1657	.1018
		(.002)	(.001)
log(price)$1_{\{w=4\}}$		.0667	.0409
		(.066)	(.043)
constant	$-.2280$	-1.3667	-.8661
	(.370)	(.001)	(.000)
Observations	86960	86960	86960
LL	-26159.114	-26059.697	-26038.224

We restricted our model on leisure markets (routes between major cities and Orlando) and business markets (routes between industry-based cities). Table 12 reports logit estimation results. We observe that the price elasticity monotonically decreases over time until the last period. This result may be explained from the perspective of price evolution and demand uncertainty. First, as illustrated in Figure 12 and Figure 13, price dispersions do not show dramatic changes over time, especially in business markets. Relatively small and stationary price distributions may cause reservation valuation to keep decreasing $v_k \geq v_{k+1}$, which result in monotonic decrease in demand elasticity. Next, we consider demand uncertainty and heterogenous consumer valuations. In leisure and business markets, consumers may be more homogenous in terms of demand uncertainty and reservation valuations. Thus, we expect demand uncertainty may have smaller effects on book-to-look in those markets.

Table 12: Regression Results: Leisure and Business Markets

Variables	Leisure	Business
log(price)	-1.5011	-1.5201
	(.000)	(.000)
log(stdev(price))	-.1860	.0547
	(.009)	(.416)
weeks from departure	1.3351	1.3360
	(.005)	(.005)
thanksgiving	.0987	-.1398
	(.740)	(.719)
log(price)$1_{\{w=0\}}$	1.0097	1.1622
	(.002)	(.015)
log(price)$1_{\{w=1\}}$	.9049	.9626
	(.012)	(.014)
log(price)$1_{\{w=2\}}$	.6819	.6928
	(.013)	(.022)
log(price)$1_{\{w=3\}}$	.4339	.4509
	(0.019)	(.035)
log(price)$1_{\{w=4\}}$	.2012	.1962
	(.049)	(.143)
constant	.1231	.9367
	(.910)	(.456)
Observations	28360	7122
LL	-5680.540	-3168.638

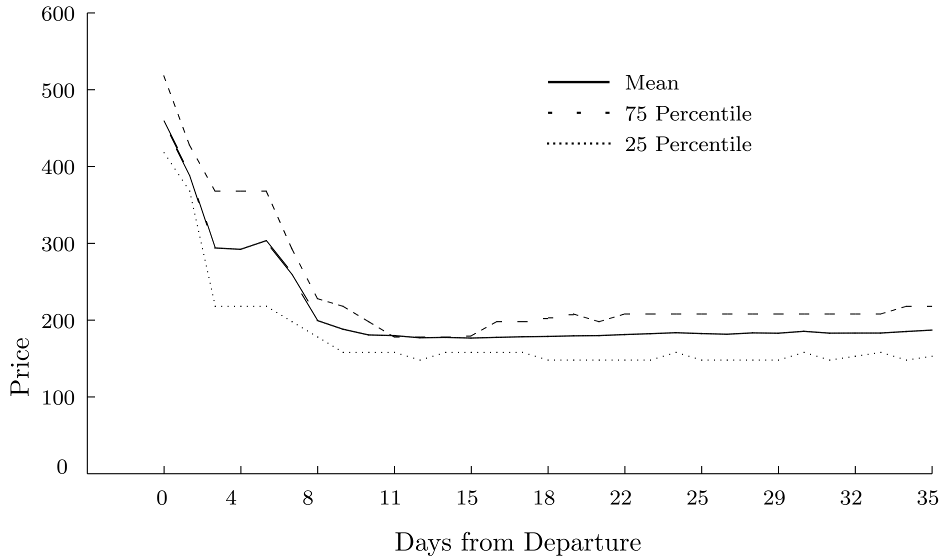

Figure 12: Evolution of mean price and price dispersion: Leisure Markets

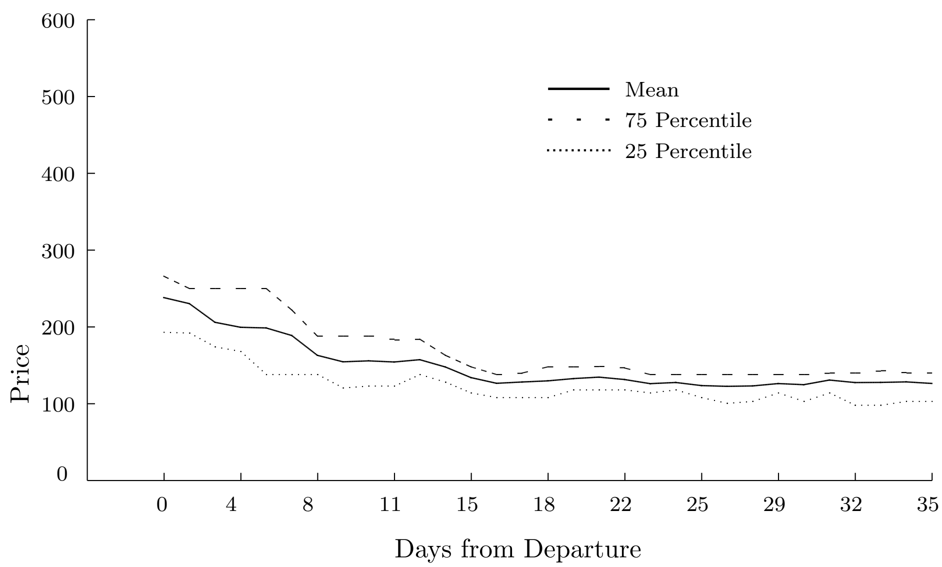

Figure 13: Evolution of mean price and price dispersion: Business Markets

CHAPTER IV

CONCLUSION

4.1 Markov Chain Model for Dynamic Search and Purchase Behaviors

The purpose of this research is to model the within-site browsing behavior of consumers both within and across visits. We focused our study on two elements: (1) key search effects on a series of consumer decisions of page request, purchase, and revisit, and (2) inter- and intra-visit dynamics of browsing behavior.

First, we observe that within-site search effects – that is, intra-visit activities (search depth and long page-view durations) as well as inter-visit exposures (repeated visits and short revisit durations) – lead to strong consumer engagement and thus increase purchase and revisit propensities. Furthermore, fit comparison between first and second order continuous markov chain models allows us to conclude that consumer choices are primarily influenced by the current search, independent of the past searches.

Next, our results indicate that consumers dynamically adjust their browsing behavior both within and across visits. There first is evidence for time constraints or learning effects. Second, our results are consistent with the presence of two distinct categories of consumer behaviors: goal-oriented and experiential consumptions that spill over multiple visits.

In this way, paper contributes to literature in online search and purchase behaviors (1) proposing the search effects within and across visits on consumer engagement (2) relating the consumer stickiness to a series of consumer decisions on page request, purchase and revisit, and (3) providing a comprehensive view on consumer decisions and decision cycles through a continuous Markov chain model that incorporates discrete choice models with the proportional hazard models.

4.2 Inter-temporal Search and Purchase Decisions of Strategic Buyers

We investigate the search and purchasing behavior of strategic buyers under a deadline. We show that time-consistent buyers follow simple behavioral rules based on future and current prices and options available. At any period, buyers buy a product if the derived utility of having exceeds a monotonically decreasing reservation valuation. We show that the trade-off between waiting and purchasing will become less and less favorable to waiting. Price elasticity should therefore drop as departure date approaches. Intuitively, as consumers become more impatient over time, they lower their reservation valuations. This consumer impatience leads to more intensive searches and higher purchase propensities.

On the other hand, the data reveal that search does increase as deadline approaches but that purchases given searches decline. To explain these mixed results on search and purchase, we extend the base model to allow for price evolution and demand uncertainty. We find that increases in mean price and price dispersion may attenuate increasing propensities for search and purchase. Intuitively, the book-to-look may go down if the tail of price distribution is dramatically getting fatter, which means the chances of getting high prices increase even faster than the reservation valuation. We also observe that demand uncertainty incorporated with heterogenous valuations may have negative effects on consumer search and purchase effects.

We demonstrate our models through a logit model estimation on a unique data set on search and purchase behavior from an internet search engine. The estimation shows that unless for the last week of purchase, demand is downward slopping. More importantly, we confirm that demand becomes more inelastic as deadline approaches. The rate of price inelasticity growth slows in the last period. The increase in mean price and price dispersion of the last period reduces the consumer utility and thus discourages purchasing, which attenuate price inelasticity. This result is consistent with decreasing book-to-look ratio in the last period.

APPENDIX A

EXAMPLES OF INTER-TEMPORAL SEARCH AND PURCHASE DECISIONS

Example 3.1.2.1. Mean-preserving spread

$$
\begin{aligned}
p_1 &= \begin{cases} \underline{p} & \text{w.p. } \beta \\ \overline{p} & \text{w.p. } 1-\beta \end{cases} \\
p_2 &= \begin{cases} \underline{p}+\epsilon & \text{w.p. } \beta \\ \overline{p}+\epsilon & \text{w.p. } 1-\beta \end{cases}
\end{aligned}
$$

where $\underline{p} < \underline{p} + \epsilon < \overline{p} < x < \overline{p} + \epsilon \leq 1$

Consumers' search cost function is defined as $c(s) = c^2$.

In the last period,

$$
\begin{aligned}
v_2 &= \max_s \left[-c(s) + sE_{p_2}[\max\{(x-p_2), 0\}]\right] \\
&= \max_s \left[-s^2 + s\big(\frac{1}{2}\beta(x-(\underline{p}-\epsilon))\big)\right] \\
&= \left(\frac{\beta(x-(\underline{p}-\epsilon))}{4}\right)^2 \\
s_2 &= \frac{\beta(x-(\underline{p}-\epsilon))}{4} \\
b_2 &= \frac{\beta(x-(\underline{p}-\epsilon))}{4}\frac{1}{2}\beta \\
bl_2 &= \frac{\beta}{2}
\end{aligned}
$$

Note that

$$\begin{aligned}(x-\underline{p}) \geq v_2 \quad &\Leftrightarrow \quad x-\underline{p}-\left(\frac{\beta\big(x-(\underline{p}+\epsilon)\big)}{2}\right)^2 \geq 0\\
&\Leftrightarrow \quad \beta^2 \leq \frac{4(x-\underline{p})}{(x-\underline{p}-\epsilon)^2}\\
&\Leftrightarrow \quad \beta \leq \frac{2\sqrt{x-\underline{p}}}{(x-\underline{p}-\epsilon)}.\end{aligned}$$

Since $1 < \frac{2\sqrt{x-\overline{p}}}{(x-\underline{p}-\epsilon)}$, we have that $x-\underline{p} \geq v_2$.

In the advance-purchase period,

$$\begin{aligned}
v_1 &= \max_s\,[-c(s)+sE_{p_1}[\max\{x-p_1, v_2\}]+(1-s)v_2]\\
&= \max_s\,\big[-c(s)+s\big(\beta(x-\underline{p})+(1-\beta)v_2\big)+(1-s)v_2\big]\\
&= \max_s\,\big[-s^2+s\big(\beta(x-\underline{p}-v_2)\big)+v_2\big]\\
&= \max_s\left[-s^2+s\beta\left((x-\underline{p})-\left(\frac{\beta(x-(\underline{p}-\epsilon))}{4}\right)^2\right)+\left(\frac{\beta(x-(\underline{p}-\epsilon))}{4}\right)^2\right]\\
&= \frac{\beta^2}{4}\left((x-\underline{p})-\left(\frac{\beta(x-(\underline{p}-\epsilon))}{4}\right)^2\right)^2+\left(\frac{\beta(x-(\underline{p}-\epsilon))}{4}\right)^2\\
s_1 &= \frac{\beta}{2}\left((x-\underline{p})-\left(\frac{\beta(x-(\underline{p}-\epsilon))}{4}\right)^2\right)\\
b_1 &= s_2P\{x-p_1>v_2\}\\
&= \frac{\beta^2}{2}\left((x-\underline{p})-\left(\frac{\beta(x-(\underline{p}-\epsilon))}{4}\right)^2\right)\\
bl_1 &= P\{x-p_1>v_2\}\\
&= \beta
\end{aligned}$$

The average book-to-look decreases over time.

Example 3.1.2.2. Increasing Mean (with constant dispersion)

$$\begin{aligned} p_1 &= \begin{cases} \underline{p} & \text{w.p. } \beta \\ \overline{p} & \text{w.p. } 1-\beta \end{cases} \\ p_2 &= \begin{cases} \underline{p}+\epsilon & \text{w.p. } \beta \\ \overline{p}+\epsilon & \text{w.p. } 1-\beta \end{cases} \end{aligned}$$

where $\underline{p} < \underline{p}+\epsilon < \overline{p} < x < \overline{p}+\epsilon \leq 1$.

Consumers' search cost function is defined as $c(s) = c^2$.

In the last period,

$$\begin{aligned} v_2 &= \max_s \left[-c(s) + sE_{p_1}[\max\{(x-p_2), 0\}]\right] \\ &= \max_s \left[-s^2 + s\beta\big(x - (\underline{p}+\epsilon)\big)\right] \\ &= \left(\frac{\beta\big(x-(\underline{p}+\epsilon)\big)}{2}\right)^2 \\ s_2 &= \frac{\beta\big(x-(\underline{p}+\epsilon)\big)}{2} \\ b_2 &= s_2\beta = \frac{\beta^2\big(x-(\underline{p}+\epsilon)\big)}{2} \\ bl_2 &= \beta. \end{aligned}$$

Note that

$$\begin{aligned} (x-\underline{p}) \geq v_2 \quad &\Leftrightarrow \quad x - \underline{p} - \left(\frac{\beta\big(x-(\underline{p}+\epsilon)\big)}{2}\right)^2 \geq 0 \\ &\Leftrightarrow \quad \beta^2 \leq \frac{4(x-\underline{p})}{(x-\underline{p}-\epsilon)^2}. \end{aligned}$$

Since $\frac{4(x-\underline{p})}{(x-\underline{p}-\epsilon)^2} \geq 1$, we have that $x - \underline{p} \geq v_2$.

Also,

$$(x-\overline{p}) \geq v_2 \quad \Leftrightarrow \quad x-\overline{p}-\left(\frac{\beta\big(x-(\underline{p}+\epsilon)\big)}{2}\right)^2 \geq 0$$
$$\Leftrightarrow \quad \beta^2 \leq \frac{4(x-\overline{p})}{(x-\underline{p}-\epsilon)^2}.$$

Now let

$$\beta' \quad = \quad \min\{\frac{2\sqrt{x-\overline{p}}}{(x-\underline{p}-\epsilon)}, 1\}.$$

In the advance-purchase period,

$$\begin{aligned}
v_1 &= \max_s \left[-c(s) + sE_{p_1}[\max\{x-p_1, v_2\}] + (1-s)v_2\right] \\
&= \begin{cases} \max_s \left[-s^2 + s\big(\beta(x-\underline{p}) + (1-\beta)(x-\overline{p})\big) + (1-s)v_2\right] & \text{if } \beta \leq \beta' \\ \max_s \left[-s^2 + s\big(\beta(x-\underline{p}) + (1-\beta)v_2\big) + (1-s)v_2\right] & \text{if } \beta > \beta' \end{cases} \\
&= \begin{cases} \max_s \left[-s^2 + s\big((x-\overline{p}-v_2) + \beta(\overline{p}-\underline{p})\big) + v_2\right] & \text{if } \beta \leq \beta' \\ \max_s \left[-s^2 + s\big(\beta(x-\underline{p}-v_2)\big) + v_2\right] & \text{if } \beta > \beta' \end{cases} \\
&= \begin{cases} \left(\frac{\big((x-\overline{p}-v_2)+\beta(\overline{p}-\underline{p})\big)}{2}\right)^2 + v_2 & \text{if } \beta \leq \beta' \\ \left(\frac{\beta(x-\underline{p}-v_2)}{2}\right)^2 + v_2 & \text{if } \beta > \beta' \end{cases} \\
s_1 &= \begin{cases} \left(\frac{\big((x-\overline{p}-v_2)+\beta(\overline{p}-\underline{p})\big)}{2}\right) & \text{if } \beta \leq \beta' \\ \left(\frac{\beta(x-\underline{p}-v_2)}{2}\right) & \text{if } \beta > \beta' \end{cases} \\
b_1 &= s_1 P(x-p_1 > v_2) \\
&= \begin{cases} \left(\frac{\big((x-\overline{p}-v_2)+\beta(\overline{p}-\underline{p})\big)}{2}\right) & \text{if } \beta \leq \beta' \\ \left(\frac{\beta^2(x-\underline{p}-v_2)}{2}\right) & \text{if } \beta > \beta' \end{cases} \\
bl_1 &= \begin{cases} 1 & \text{if } \beta \leq \beta' \\ \beta & \text{if } \beta > \beta' \end{cases}
\end{aligned}$$

The book-to-look decreases over time.

We also may observe $s_1 \leq s_2$. For example, for $x = 0.5, \underline{p} = 0.1, \overline{p} = 0.49, \epsilon = 0.02$ and $\beta = 0.9$, $s_1 = 0.165 < s_2 = 0.171$.

Example 3.1.3. Demand uncertainty and heterogeneous consumer valuation (with stationary price distributions)

Suppose that

$$p_1 = p_2 = \begin{cases} \underline{p} & \text{w.p. } \beta_1 \\ p' & \text{w.p. } \beta_2 \\ p'' & \text{w.p. } \beta_3 \\ \overline{p} & \text{w.p. } 1-(\beta_1+\beta_2+\beta_3) \end{cases}$$

There are two types of consumers, high (x_h) and low(x_l) valuation consumers, where

$$\underline{p} < p' < x_l < p'' < \overline{p} < x_h \leq 1.$$

In the last period, the book-to-look of low value consumers is $bl_2^l = P\{x_l - p_2 > 0\} = \beta_1 + \beta_2$. In the meantime, the book-to-look of high value consumers is $bl_2^h = P\{x_h - p_2 > 0\} = 1$.

Specifically,

$$\begin{aligned} v_2^l &= \max_s \left[-c(s) + sE_{p_2}[\max\{(x_l - p_2), 0\}]\right] \\ &= \max_s \left[-s^2 + s\big(\beta_1(x_l - \underline{p}) + \beta_2 \cdot (x_l - p')\big)\right] \\ &= \left(\frac{\beta_1(x_l - \underline{p}) + \beta_2 \cdot (x_l - p')}{2}\right)^2 \\ s_2^l &= \frac{\beta_1(x_l - \underline{p}) + \beta_2 \cdot (x_l - p')}{2} \\ b_2^l &= \frac{\beta_1(x_l - \underline{p}) + \beta_2 \cdot (x_l - p')}{2}(\beta_1 + \beta_2) \\ bl_2^l &= \beta_1 + \beta_2 \end{aligned}$$

and

$$
\begin{aligned}
v_2^h &= \max_s \left[-c(s) + sE_{p_1}[\max\{(x_h - p_2), 0\}]\right] \\
&= \max_s \left[-s^2 + s\big(\beta_1(x_l - \underline{p}) + \beta_2 \cdot (x_l - p') + \beta_3(x_l - p'') + (1-\beta_1-\beta_2-\beta_3)(x_l - \overline{p})\big)\right] \\
&= \left(\frac{\beta_1(x_l - \underline{p}) + \beta_2 \cdot (x_l - p') + \beta_3(x_l - p'') + (1-\beta_1-\beta_2-\beta_3)(x_l - \overline{p})}{2}\right)^2 \\
s_2^h &= \frac{\beta_1(x_l - \underline{p}) + \beta_2 \cdot (x_l - p') + \beta_3(x_l - p'') + (1-\beta_1-\beta_2-\beta_3)(x_l - \overline{p})}{2} \\
b_2^h &= \frac{\beta_1(x_l - \underline{p}) + \beta_2 \cdot (x_l - p') + \beta_3(x_l - p'') + (1-\beta_1-\beta_2-\beta_3)(x_l - \overline{p})}{2} \\
bl_2^h &= 1.
\end{aligned}
$$

Note that $x_l - \underline{p} \geq v_2^l$. Assume that $x_l - p' \leq v_2^l$.

In the advance-purchase period,

$$
\begin{aligned}
v_1^l &= \max_s \left[-c(s) + sE_{p_1}[\max\{\alpha_l x_l - p_1, \alpha_l v_2^l\}] + (1-s)\alpha_l v_2^l\right] \\
&= \max_s \left[-c(s) + s\big(\beta_1(x_l - \underline{p}) + (1-\beta_1)v_2^l\big) + (1-s)v_2^l\right] \\
&= \max_s \left[-s^2 + s\beta_1\big(x_l - \underline{p} - v_2^l\big) + v_2^l\right] \\
&= \left(\frac{\beta_1\big(x_l - \underline{p} - v_2^l\big)}{2}\right)^2 + v_2^l \\
s_1^l &= \frac{\beta_1\big(x_l - \underline{p} - v_2^l\big)}{2} \\
b_1^l &= \frac{\beta_1\big(x_l - \underline{p} - v_2^l\big)}{2}\beta_1 \\
bl_1^l &= P\{x_l - p_1 > v_2^l\} = \beta_1
\end{aligned}
$$

Assume that $\alpha_h x_h - p'' > v_2^h$ and $\alpha_h x_h - \overline{p} < v_2^h$.

$$
\begin{aligned}
v_1^h &= \max_s \left[-c(s) + sE_{p_1}[\max\{\alpha_h x_h - p_1, \alpha_h v_2^h\}] + (1-s)\alpha_h v_2^h\right] \\
&= \max_s \Big[-s^2 + s\big(\beta_1(\alpha_h x_h - \underline{p}) + \beta_2(\alpha_h x_h - p') + \beta_3(\alpha_h x_h - p'') + (1-\beta_1-\beta_2-\beta_3)\alpha_h v_2^h\big) + (\\
&= \max_s \Big[-s^2 + s\big(\beta_1(\alpha_h x_h - \underline{p} - \alpha_h v_2^h) + \beta_2(\alpha_h x_h - p' - \alpha_h v_2^h) + \beta_3(\alpha_h x_h - p'' - \alpha_h v_2^h)\big) + \alpha_h v \\
&= \left(\frac{\beta_1(\alpha_h x_h - \underline{p} - \alpha_h v_2^h) + \beta_2(\alpha_h x_h - p' - \alpha_h v_2^h) + \beta_3(\alpha_h x_h - p'' - \alpha_h v_2^h)}{2}\right)^2 + \alpha_h v_2^h \\
s_1^h &= \frac{\beta_1(\alpha_h x_h - \underline{p} - \alpha_h v_2^h) + \beta_2(\alpha_h x_h - p' - \alpha_h v_2^h) + \beta_3(\alpha_h x_h - p'' - \alpha_h v_2^h)}{2} \\
b_1^h &= s_1^h P(\alpha_h x_h - p_1 > \alpha_h v_2^h) \\
&= \frac{\beta_1(\alpha_h x_h - \underline{p} - \alpha_h v_2^h) + \beta_2(\alpha_h x_h - p' - \alpha_h v_2^h) + \beta_3(\alpha_h x_h - p'' - \alpha_h v_2^h)}{2}(\beta_1+\beta_2+\beta_3) \\
bl_1^h &= P(\alpha_h x_h - p_1 > \alpha_h v_2^h) \\
&= \beta_1 + \beta_2 + \beta_3
\end{aligned}
$$

REFERENCES

[1] Aaron, K., "SITA survey: Choice is key to encouraging self-service," *Air Transport World* , pp. 76–80, 2007.

[2] Allenby, G. M., Leone, R. P., and Jen, L., "A Dynamic Model of Purchase Timing With Application to Direct Marketing," *Journal of the American Statistical Association*, vol. 94 , pp. 365-374, 1999.

[3] Buchanan, B.S. and Morrison, D.G., "A Stochastic Model of List Falloff with Implications for Repeat Mailings", *Journal of Direct Marketing*, vol. 2(3), pp. 7–15, 1988.

[4] Bucklin, Randolf E., Catarina Sismeiro, "A model of Web site browsing behavior estimated on clickstream data", *J. Marketing Res.*, vol. 40(3), pp. 249–267, 2003.

[5] Chatterjee, Patrali, Donna L. Hoffman, Thomas P. Novak, "Modeling the Clickstream: Implications for Web-based Advertising Efforts", *Marketing Science*, vol. 22(4), pp. 520–541, 2003.

[6] Dana, James, "Advanced Purchase discounts and Price Discrimination in Competitive Markets," *Journal of Political Economy* , pp. 395–422, April 1998.

[7] Gallego, G. and G.J. van Ryzin, "Optimal Dynamic Pricing of Inventories with Stochastic Demand Over Finite Horizons," *Management Science*, vol. 40 , pp. 999–1020, 1993.

[8] Harteveldt, Henry. H., Charles P. Wilson & Carrie A. Johnson, "Why leisure travelers book at their favorite sites," *Forrester Trends Document*, pp. 1–2, 2004.

[9] Harteveldt, Henry. H., "Leisure travel data overview," *Forrester Research: Data Overview*, 2006.

[10] Hoffman, Donna L. and Thomas P. Novak, "Marketing in hypermedia computer-mediated environments: Conceptual foundations", *J. Marketing*, vol. 60, pp. 50–68, July 1996.

[11] Horner, Johannes and Larry Samuelson, "Managing Strategic Buyers," *Working Paper*, 2008.

[12] Jain, D. C., and Vilcassim, N. J., "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach", *Marketing Science*, vol. 10, pp. 1–23, 1991.

[13] Johnson, Eric J., Steven Bellman, and Jerry Lohse, "Cognitive Lock-In and the Power Law of Practice", *Journal of Marketing*, vol. 67, pp. 62–75, April 2003.

[14] Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman & Gerald L. Lohse, "On the depth and dynamics of online search behavior", *Management Science*, vol. 50(3), pp. 299–308, 2004.

[15] Mandel, Naomi and Eric J. Johnson, "When Web Pages lnfluence Choice: Effects of Visual Primes on Experts and Novices", *Journal of Consumer Research*, vol. 29(2), pp. 235–245, 2002.

[16] Moe, Wendy W., "Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream", *Journal of Consumer Psychology*, vol. 13(1 & 2), pp. 29–39, 2003.

[17] Moe, Wendy W. and Peter S. Fader, "Dynamic conversion behavior at e-commerce sites", *Management Science*, vol. 50(3), pp. 326–335, 2004.

[18] Montgomery, Alan L., Shibo Li, Kannan Srinivasan & John C. Liechty, "Modeling online browsing and path analysis using clickstream data", *Marketing Science*, vol. 23(4), pp. 579–595, 2004.

[19] PhoCusWright, "The PhoCusWright Consumer Travel Trends Survey," *PhoCusWright downloaded from http:// store.phocuswright.com*, 2008.

[20] Sargent, T., *Dynamic Macroeconomic Theory*, Harvard University Press, 1987.

[21] Seetharaman, P. B. and Pradeep K. Chintagunta, "The Proportional Hazard Model for Purchase Timing: A Comparison of Alternative Specifications", *Journal of Business & Economic Statistics*, vol. 21(3), pp. 369–382, 2003.

[22] Smith, Barry, Ross Darrow, John Elieson, Dirk Guenther & B. Venkateshwara Rao, "Travelocity becomes a travel retailer", *Interfaces*, vol. 37(1), pp. 68–81, 2007.

[23] Wooldridge, Jeffrey, *Introductory Econometrics: A Modern Approach*, 3 edition. South-Western College Pub., 2005.

[24] Zauberman, Gal, "The Intertemporal Dynamics of Consumer Lock-In", *Journal of Consumer Research*, vol. 30, pp. 405–419, December 2003.